Genericity in Linear Algebra and Analysis with Application to Optimization

October 4, 2019

Georg Still
University of Twente

Contents

CHAPTER 1

Introduction and notation

1.1. Introduction

This first section aims to provide an introduction into the topic of the booklet. We start with an illustrative example to motivate the genericity concept.

Consider the problem of solving linear equations,

$$(1.1) \qquad (P): \qquad Ax = b$$

with a matrix $A \in \mathbb{R}^{n\times n}$, and a vector $b \in \mathbb{R}^n, \ \ n \in \mathbb{N}$. It is well-known that this equation has a unique solution for all b if and only the following property is met:

(1.2) **Property.** A is nonsingular, or equivalently, $\det(A) \neq 0$.

So we could ask the following

Question. Can we expect that for problem (P) in the "normal" (or "generic") case (to be defined lateron) the "nice" property (1.2) holds?

To formulate this question properly we have to define what we mean by a problem and what we mean by "generic case".
For our example problem (P) and the property in (1.2) a problem instance is given by a matrix $A \in \mathbb{R}^{n\times n}$. If we specify the "size" n of A, we can consider the whole family $\mathcal{P}$ (problem class $\mathcal{P}$) of problems (P), and we can identify this problem class with the set of all problem instances A:

$$(1.3) \qquad \mathcal{P} = \{A \mid A \in \mathbb{R}^{n\times n}\} \ \ \equiv \ \ \mathbb{R}^{n\times n} \ .$$

For topological results, e.g., if we speak of open sets, we need to specify some topology on $\mathbb{R}^{n\times n}$. Here, we chose the topology related with the (Frobenius-) norm $\|A\| = \sqrt{\sum_{i,j} a_{ij}^2}$, for a matrix $A = (a_{ij})_{i,j=1,\dots,n}$.

DEFINITION 1.1. [problem, problem instance, problem class, problem set (data)]
We denote, e.g., *by (P), a mathematical problem of a special type and are interested in a specific property for (P). Take for example (P) in* (1.1) *and property* (1.2)*. This problem and the property depends on the problem data (A in our example). If we specify the data, then we speak of an instance Q of problem (P) ($Q = A$ in our example).*

By a class $\mathcal{P}$ of problems (P) we mean a family of problem instances Q of a specified size. In this report we assume that the problem instances Q are elements of some linear real vector space $\mathcal{L}$ endowed with some norm $\| \|_{\mathcal{L}}$, or some other topology.
So, the problem class $\mathcal{P}$ can be identified with the set of instances Q (problem set in short):

$$\mathcal{P} = \{\text{a subset of instances } Q \text{ in } \mathcal{L}\}$$

We often tacitly assume that the problem size (n in the example) is fixed and simply use the notion problem to denote the problem class.

Coming back to the problem (P) in (1.1), the property in (1.2), and the problem class $\mathcal{P}$ in (1.3), we have to make precise what we mean if we say that property (1.2) holds generically in $\mathcal{P}$. Our first definition is restricted to the case that the problem set (data set) $\mathcal{P}$ is a subset of a space $\mathbb{R}^n$. We denote the Lebesgue measure in $\mathbb{R}^n$ by μ.

DEFINITION 1.2. [generic set, weakly generic set, generic property]

(**a**) *Let be given an open set $S \subset \mathbb{R}^n$. A subset $S_0 \subset S$ is called a generic set in S if*

(1) the complement $S \setminus S_0$ of S_0 has Lebesgue measure zero, i.e., $\mu(S \setminus S_0) = 0$, *and*

(2) *the set S_0 is open in $\mathbb{R}^n$.*

The set S_0 is called weakly generic if only the first condition (1) holds. We also simply call a set $S_r \subset S$ generic in S if S_r contains such a generic subset S_0 in S.

(**b**) *Let us consider a problem class $\mathcal{P}$ of a problem (P), where $\mathcal{P} \subset \mathbb{R}^n$ is open. A property is called to be generic in the problem class $\mathcal{P}$ if there is a generic subset $\mathcal{P}_r \subset \mathcal{P}$, i.e.,*

(1) $\mathcal{P} \setminus \mathcal{P}_r$ has Lebesgue measure zero in $\mathcal{P}$, $\mu(\mathcal{P} \setminus \mathcal{P}_r) = 0$ and

(2) the set $\mathcal{P}_r$ is open in $\mathbb{R}^n$,

such that the property holds for all instances $Q \in \mathcal{P}_r$. We then also say that the property is generically fulfilled in $\mathcal{P}$.
The property is said to be weakly generic if for $\mathcal{P}_r$ only the first condition (1) holds.

REMARK 1.1. The condition (1) in the definition above says that the property holds for almost all instances from $\mathcal{P}$. The second part (2) implies that the property is stable at any instance $\overline{Q} \in \mathcal{P}_r$. This condition (2) is particularly important in practical applications. It means that near $\overline{Q} \in \mathcal{P}_r$ the ("nice") property is stable wrt. (sufficiently) small perturbations of the instance (data) $\overline{Q}$.

REMARK 1.2. For the definition of a generic set in $\mathbb{R}$ (or $\mathbb{R}^n$) we have chosen the Lebesgue measure. Other choices are possible. We have chosen this measure since it is common in analysis and the definition of a set of Lebesgue measure zero only involves basic notions (see Definition 2.1).
If we consider another (finite) measure λ in $\mathbb{R}$ which is defined by a distribution function $F(x) = \int_{-\infty}^{x} \delta \, d\mu$ with nonnegative density function $\delta : \mathbb{R} \to \mathbb{R}_+$ such that the Lebesgue

integral $\int_{-\infty}^{\infty} \delta \; d\mu$ exists, i.e., $\lambda(S) = \int_S \delta \; d\mu$ for any Lebesgue measurable set $S \subset \mathbb{R}$, then $\mu(S) = 0$ implies $\lambda(S) = 0$ (see e.g., [34, 11.23 Remark]). So, for such measures λ, genericity wrt. the Lebesgue measure implies genericity wrt. λ measure. The Lebesgue measure corresponds to a uniform distribution given by a "constant" density function.

Let us reconsider property (1.2) for problem (P) in (1.1), with given $n \in \mathbb{N}$, and corresponding problem set $\mathcal{P} \equiv \mathbb{R}^{n\times n}$. To show that the property is generic, we have to show that the subset $\mathcal{P}_r$ of $\mathcal{P}$ given by

$$\mathcal{P}_r = \{A \in \mathcal{P} \equiv \mathbb{R}^{n\times n} \mid \det(A) \neq 0\} \tag{1.4}$$

is generic. To do so, we apply the basis result in Theorem 3.1. According to formula (7.1) the function $p(A) = \det(A)$ defines a polynomial function on $\mathbb{R}^{n\times n}$, which by $det(I_n) = 1$, is not the zero function. So, using Theorem 3.1, the set $\mathcal{P} \setminus \mathcal{P}_r = \{A \in \mathcal{P} \mid \det(A) = 0\}$ has Lebesgue measure zero in $\mathcal{P}$. To show the openness condition for $\mathcal{P}_r$ we conclude from (7.1) that $\det(A)$ is a continuous function of A. So if $\det(\overline{A}) \neq 0$ holds, i.e., $\overline{A} \in \mathcal{P}_r$, then there must exist some $\varepsilon > 0$ such that $\det(A) \neq 0$ is true for all $A \in \mathcal{P}$ satisfying $\|A - \overline{A}\| < \varepsilon$. Consequently, $\mathcal{P}_r$ is open and we have obtained our first genericity result.

THEOREM 1.1. *The subset $\mathcal{P}_r$ in* (1.4) *is generic in $\mathcal{P}$ and thus the property that $Ax = b$ has a unique solution for all $b \in \mathbb{R}^n$ is generic in $\mathcal{P}$.*

This genericity result can equivalently be formulated as follows: The property that (for all $b \in \mathbb{R}^n$) the linear mapping $F(x) := Ax - b$ has a nonsingular Jacobian $\nabla F(x) = A$ is generic in $\mathcal{P} = \{A \mid A \in \mathbb{R}^{n\times n}\}$.

Nonlinear case:

If we consider problems where general nonlinear functions are involved, things become more complicated. Let us look at the problem of solving the system of n nonlinear equations in n variables:

$$(P): \qquad F(x) = 0 \; , \tag{1.5}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a function from $C^1(\mathbb{R}^n, \mathbb{R}^n)$. The problem set would then be

$$\mathcal{P} = \{F \mid F \in C^1(\mathbb{R}^n, \mathbb{R}^n)\} \equiv [C^1(\mathbb{R}^n, \mathbb{R})]^n \; ,$$

where the space $C^1(\mathbb{R}^n, \mathbb{R})$ is endowed with some topology, say a C^1_t-topology. How could we define genericity wrt. such a function space and what could be a property similar to (1.2)? In contrast to the linear case $F(x) := Ax - b$, for the system (1.5) of nonlinear equations, even in the "nice" situation we may have no solution, or many different solutions (even infinitely many, see Figure 1.1). So we only can expect that "generically" the following "nice" property holds:

Property 1. At each solution $\overline{x} \in \mathbb{R}^n$ of $F(x) = 0$ the Jacobian

$$\nabla F(\overline{x}) \qquad \text{is nonsingular .} \tag{1.6}$$

Later, by applying the Implicit Function Theorem 7.2, we will see that condition (1.6) implies that any solution $\overline{x}$ of $F(x) = 0$ is locally unique (isolated).

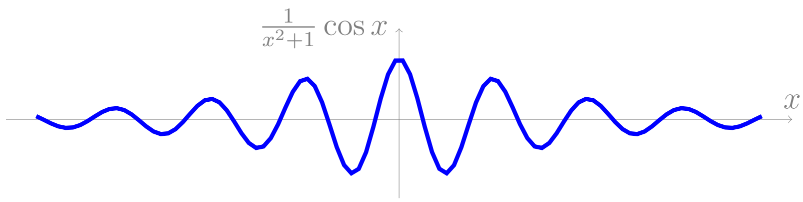

FIGURE 1.1. Infinitely many solutions of $F(x) = \frac{1}{x^2+1} \cos x = 0$

In Section 4.4 the following genericity result will be proven. In this theorem the set $C^1(\mathbb{R}^n, \mathbb{R}^n)$ is endowed with the so-called (strong) C^1_s-topology (see Section 2.2 for details).

THEOREM 1.2. *The subset $\mathcal{P} = C^1(\mathbb{R}^n, \mathbb{R}^n)$ contains a generic subset $\mathcal{P}_r$ (generic wrt. the C^1_s-topology), such that Property 1 is satisfied for any $F \in \mathcal{P}_r$.*

As we shall see in Section 4.4, this results is a basis result for the Newton method for solving systems (1.5). In fact this theorem says that generically the assumptions for the (local) quadratic convergence of the Newton method is met. But before proving such a result in Section 4.4 we first have to answer the following questions:

- How to define a generic set, *e.g.*, in $C^1(\mathbb{R}^n, \mathbb{R})$?
  Rough answer: We will chose the C^1_s-topology for $C^1(\mathbb{R}^n, \mathbb{R})$ (see Section 2.2).

- Which techniques are needed to prove the genericity results.
  Rough answer: We will apply techniques from Differential Geometry or Differential Topology (cf., Section 4.1).

REMARK 1.3. Let us note that the stronger property:

$$\textit{Property 2:} \qquad \nabla F(x) \quad \text{is nonsingular for all } x \in \mathbb{R}^n \ ,$$

cannot be expected to hold generically in $C^1(\mathbb{R}^n, \mathbb{R}^n)$. Take the example $F \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ given by (cf., Figure 1.2)

$$F(x) = \sin x \quad \text{with} \quad \nabla F(x) = \cos x \ .$$

Here, Property 1 is fulfilled: at each solution $\overline{x}_k = k\pi, k \in \mathbb{Z}$, of $F(x) = 0$, the condition

$$\nabla F(\overline{x}_k) = \cos k\pi = (-1)^k \neq 0$$

is valid. However, Property 2 does not hold and we cannot expect that by small (differentiable) perturbations of F we obtain some $\hat{F}(x) \approx \sin x$, $\hat{F} \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, such that $\nabla\hat{F}(x) \approx \cos x$ satisfies

$$\nabla\hat{F}(x) \neq 0 \quad \text{for all } x \in \mathbb{R}^n.$$

<u>A General warning:</u>

Throughout this report many genericity results in linear algebra, analysis, and optimization

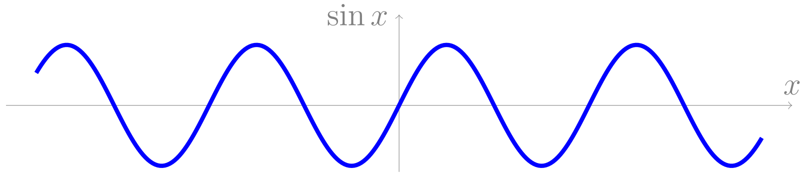

FIGURE 1.2. Function $F(x) = \sin x$ satisfies $F'(\overline{x}) \neq 0$ at all solutions $\overline{x}$ of $F(\overline{x}) = 0$.

will be presented. We however have to give a warning. The genericity results always concern a specific class of problems (with a certain structure). The point is, that a genericity result for a given problem set cannot be directly transferred to a subclass (with a specific substructur) of this problem set. Unfortunately each subclass of a problem class which consists of data with a special structure requires a new (independent) genericity analysis. In what follows we will regularly come back to this fact. For example, the genericity results for general real $(n \times n)$-matrices need no more be true for matrices with a special structure such as band matrices (see, Section 5.3 for a counterexample). Also the genericity results for (general) optimization problems (non-parametric in Section 4.5, or parametric in Section 6.5) are not valid for the subclass of quadratic or linear programs (see, Section 6.6).

A comment on the practical relevance of genericity results:

A genericity result on a problem class is a statement on the structure shared by almost all problems of the class (see also Remark 1.1). What is the relevance of such results on the generic behavior of a specific class of mathematical problems? Beyond their mathematical value they are important for all solution methods for solving the problems numerically. A specific solution method typically requires some regularity conditions to be satisfied for the problem at hand. So, from a practical viewpoint, it is important that the required regularity conditions are generically fulfilled within the problem class. Otherwise it is likely that the method fails for most of the problems in the class.Therefore it is crucial that a solution method is designed in such a way that it is able to cope numerically with all situations which can occur generically. Ideally the algorithm should also be able to detect situations which are not generic.

1.2. Summary and comments on further reading

We start with a short summary of the contents of this report. In Chapter 2 we present the definitions and basic properties of generic sets in the space $\mathbb{R}^n$ and generic sets in function spaces $C^k(\mathbb{R}^n, \mathbb{R})$. We further provide an introduction into algebraic- and semi-algebraic sets. Chapter 3 performs a genericity analysis for (non-parametric) linear programs and Chapter 4 deals with genericity results for non-parametric problems in analysis, such as systems of nonlinear equations as well as unconstrained/constrained programs.

Chapter 5 provides the basis techniques needed for the analysis of parametric problems in Chapter 6. In the latter, genericity results for parametric families of matrices and parametric families of optimization problems are presented. In particular the structure of one-parametric general programs as well as one-parametric quadratic and linear programming problems are studied in detail.
The genericity results in this report are based on (Parametric) Sard Theorems and techniques from René Thom's transversality theory in Differential Topology.
We emphasize that much of the ideas and material discussed in this report are taken from the landmark book [21] and its predecessors [19, 20]. These investigations on the generic behavior of unconstrained/constrained optimization problems have inspired many researchers to study the generic properties of other classes of mathematical programs. We only sum up some of these studies:
In [41], genericity results for (non-parametric) semi-infinite optimization problems

$$(P_{SIP}): \quad \min_{x \in \mathbb{R}^n} \ f(x) \quad \text{s.t.} \quad g(x,y) \leq 0 \ \forall y \in Y \ ,$$

with $Y \subset \mathbb{R}^m$, a compact index set, are presented. In [35] and [24], one-parametric semi-infinite problems are studied, leading in particular to a genericity theorem of the 8 Types. The generic behavior of so-called generalized semi-infinite programs is investigated in [15].
For genericity results for programs with complementarity constraints of the form,

$$(P_{CC}): \quad \min_{x \in \mathbb{R}^n} \ f(x) \quad \text{s.t.} \quad \begin{array}{rl} g_j(x) \leq 0 \ , & j = 1, \ldots, m \\ s_i(x) \cdot r_i(x) = 0 \ , & i = 1, \ldots, l \\ s_i(x), \ r_i(x) \geq 0 \ , & i = 1, \ldots, l \end{array} \quad ,$$

we refer to [38], [6] for the non-parametric case and to [7] for one-parametric (P_{CC}) programs. Also the generic behavior of bilevel problems of the following form have been studied in [8]:

$$(P_{BL}): \quad \min_{x,y} \ f(x,y) \qquad \text{s.t.} \quad g_j(x,y) \leq 0 \ , \ j = 1, \ldots, m \ ,$$
$$\text{and s.t. } y \text{ solves the program } Q(x) \ ,$$

where $Q(x)$ is the so-called lower level problem:

$$Q(x): \quad \min_{y} \ h(x,y) \quad \text{s.t.} \quad v_i(x,y) \leq 0 \ , \ i = 1, \ldots, l \ .$$

Genericity results for linear bilevel problems have been proven before in [37].
A summary of the generic behavior of (non-parametric) linear conic programs is to be found in [12]. In [1] the special (non-parametric) maximization problem of homogeneous polynomials on the unit simplex has been investigated from a generic viewpoint.

We emphasize that the present report only considers mathematical problems with unknowns in a space $\mathbb{R}^n$. But there are also genericity results for many problems where the unknowns are functions $u(x)$ in some Banach space such as problems where partial differential equations are involved. Take for example the eigenvalue problem with the Laplacian operator $\Delta u(x_1, x_2) = \frac{\partial^2}{\partial x_1^2} u(x_1, x_2) + \frac{\partial^2}{\partial x_2^2} u(x_1, x_2)$: find a function $u : G \to \mathbb{R}$, $u \neq 0$, and a value $\lambda \in \mathbb{R}$ such that

$$(E): \quad \Delta u(x_1, x_2) + \lambda u(x_1, x_2) = 0 \quad \forall\, (x_1, x_2) \in G \text{ and } u(x_1, x_2) = 0\ \forall (x_1, x_2) \in \operatorname{bd} G,$$

where G is some given compact set in $\mathbb{R}^2$. It has, e.g., been proven in [2] that "generically" the eigenvalue problem (E) has simple eigenvalues (see also the results for algebraic eigenvalues in the Sections 4.2, 6.2.2).

1.3. Notation

To facilitate the reading we present a list of notation used in this booklet.

$\mathbb{N}$	$\mathbb{N} = \{0, 1, 2, \ldots\}$ or depending on the context $\mathbb{N} = \{1, 2, \ldots\}$
e_j	in $\mathbb{R}^n$, e_j denote the standard unit vectors
$x \in \mathbb{R}^n$	column vector with n components $x_i, i = 1, \ldots, n$
$x^T \in \mathbb{R}^n$	transposed of x, row vector $x^T = (x_1, \ldots, x_n)$
x_{J_0}	for $x \in \mathbb{R}^n$ and a subset $J_0 \subset \{1, \ldots, n\}$, it is the (partial) vector $x_{J_0} = (x_j,\ j \in J_0)^T$
$\\|x\\|$	the Euclidean norm of $x \in \mathbb{R}^n$, $\\|x\\| = \sqrt{x_1^2 + \ldots + x_n^2}$
$\\|x\\|_\infty$	the maximum norm of $x \in \mathbb{R}^n$, $\\|x\\|_\infty = \max_{1 \le i \le n} \|x_i\|$
$B(\overline{x}, \varepsilon)$	for $\overline{x} \in \mathbb{R}^n$, $\varepsilon > 0$, it is the neigborhood $B(\overline{x}, \varepsilon) = \{x \in \mathbb{R}^n \mid \\|x - \overline{x}\\| < \varepsilon\}$
$\mathbb{R}^n_+$	set of $x \in \mathbb{R}^n$ with components $x_i \ge 0, i = 1, \ldots, n$
$A = (a_{ij})$	$A \in \mathbb{R}^{n \times m}$, a real $(n \times m)$-matrix, with element a_{ij} in row i, column j
$A_{i\cdot}$	is the ith row of A
$A_{\cdot j}$	is the jth column of A
A_{J_0}	for an $(n \times m)$-matrix A and a subset $J_0 \subset \{1, \ldots, n\}$, A_{J_0} is the (partial) matrix $A_{J_0} = \begin{pmatrix} A_{j\cdot} \\ \vdots \end{pmatrix}_{j \in J_0}$
0	0 vector or 0 matrix of appropriate dimension
$\\|A\\|$	Frobenius norm, $\\|A\\| = \sqrt{\sum_{ij} a_{ij}^2}$. This matrix norm satisfies: $\\|Ax\\| \le \\|A\\|\\|x\\|$
M^n	set of real $(n \times n)$-matrices
S^n	set of real, symmetric $(n \times n)$-matrices
$M^{n,m}$	set of real $(n \times m)$-matrices with n rows and m columns
$C^0(\mathbb{R}^n, \mathbb{R})$	set of real valued functions f, continuous on $\mathbb{R}^n$. We often simply write $f \in C^0$
$C^k(\mathbb{R}^n, \mathbb{R})$	set of functions f, k-times continuously differentiable on $\mathbb{R}^n$ $(k \ge 1)$. We often simply write $f \in C^k$
$C^\infty(\mathbb{R}^n, \mathbb{R})$	set of functions$f$, infinitely many times continuously differentiable on $\mathbb{R}^n$. We often write $f \in C^\infty$
C^k_s	strong topology on $C^k(\mathbb{R}^n, \mathbb{R})$, see Definition 2.2
$\nabla f(x)$	row vector, gradient of $f \in C^1(\mathbb{R}^n, \mathbb{R})$, $\nabla f(x) = (\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n})$
$\nabla^T f(x)$	or $\nabla f(x)^T$, column vector, gradient of f, $\nabla^T f(x) = (\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n})^T$
$\nabla^2 f(x)$	Hessian of $f \in C^2(\mathbb{R}^n, \mathbb{R})$, $\nabla^2 f(x) = \left(\frac{\partial^2}{\partial x_j \partial x_i} f(x)\right)_{i,j}$
$\nabla_x f(x, t)$	partial derivatives wrt. x, $\nabla_x f(x, t) = (\frac{\partial f(x,t)}{\partial x_1}, \ldots, \frac{\partial f(x,t)}{\partial x_n})$
$\nabla^2_x f(x, t)$	Hessian wrt. x of $f(x, t)$
$\nabla \nabla^T_x f(x, t)$	or $\nabla_{(x,t)} \nabla^T_x f(x, t)$, for $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, it is a $n \times (n + p)$-matrix
$\\|f\\|_{k,C}$	norm on $C^k(\mathbb{R}^n, \mathbb{R})$ wrt. $C \subset \mathbb{R}^n$, C compact, (see (2.6)) $\\|f\\|_{k,C} := \max_{\|\alpha\| \le k}\ \max_{x \in C}\ \|\partial^\alpha f(x)\|$

$x_l \to \overline{x}$	given a sequence $x_l \in \mathbb{R}^n,\ l \in \mathbb{N}$, by $x_l \to \overline{x}$ we mean $\lim_{l\to\infty} x_l = \overline{x}$ or $\lim_{l\to\infty} \|x_l - \overline{x}\| = 0$
$J_{(\overline{x},\overline{t})}$	active index set at $(\overline{x},\overline{t})$, $J_{(\overline{x},\overline{t})} = \{j \in J \mid g_j(\overline{x},\overline{t}) = 0\}$
$L(x,t,\mu)$	Lagrangean function (locally at $(\overline{x},\overline{t})$): $L(x,t,\mu) = f(x,t) + \sum_{j\in J_{(\overline{x},\overline{t})}} \mu_j g_j(x,t)$
$C_{(\overline{x},\overline{t})}$	the cone of critical directions, see Lemma 4.5
$T_{(\overline{x},\overline{t})}$	tangent space of critical directions, see (4.21)
$N_{\overline{x}}M$	normal space of manifold M at $\overline{x} \in M$, see Def. 5.2
$T_{\overline{x}}M$	tangent space of manifold M at $\overline{x} \in M$, see Def. 5.2
$P(t)$	parametric program see (6.30)
$\mathcal{F}(t)$	feasible set of $P(t)$
$v(t)$	value function of $P(t)$
$S(t)$	set of (global) minimizers of $P(t)$
cl S	closure of the set S
int S	interior of the set S
$M_1 \overline{\pitchfork} M_2$	manifolds M_1, M_2 intersect transversally, see Def. 5.6
$f \overline{\pitchfork} M$	function f meets manifold M transversally, see Def. 5.7

CHAPTER 2

Basis Theory

After the definition of generic sets in $\mathbb{R}^n$ (cf., Definition 1.1), the present chapter gives some facts on properties of generic sets in $\mathbb{R}^n$ and introduces the genericity concept in function spaces. Further, a survey of main properties of algebraic and semi-algebraic sets is added.

2.1. Properties of generic sets in $\mathbb{R}^n$

In this section we recall some basic facts. The set

$$q = \{x \in \mathbb{R}^n \mid |x_i - \overline{x}_i| \leq \frac{\alpha}{2}\} \quad \text{with some } \overline{x} \in \mathbb{R}^n \ , \ \text{ and } \alpha > 0 \ ,$$

is called a (closed) cube in $\mathbb{R}^n$ of volume $\mu(q) := \alpha^n$.

DEFINITION 2.1. *A subset $S \subset \mathbb{R}^n$ is said to be of Lebesgue measure zero (notation $\mu(S) = 0$), if for every $\varepsilon > 0$ there is a countable family $\{q_\nu\}_{\nu \in I}$ of cubes $q_\nu \subset \mathbb{R}^n$, with volumes $\mu(q_\nu)$, $\nu \in I$ $(I \subset \mathbb{N})$, such that*

$$S \subset \bigcup_{\nu \in I} q_\nu \qquad \textit{and} \qquad \Big[\mu\big(\cup_{\nu \in I} q_\nu\big) \leq \Big] \sum_{\nu \in I} \mu(q_\nu) < \varepsilon \ .$$

We then say that $\{q_\nu\}_{\nu \in I}$ covers S.

LEMMA 2.1. *Let $\{S_i\}_{i \in \mathbb{N}}$ be a countable family of sets $S_i \subset \mathbb{R}^n$ with $\mu(S_i) = 0, \ i \in \mathbb{N}$. Then*

$$\mu\big(\cup_{i \in \mathbb{N}} S_i\big) = 0.$$

Proof. We can choose $\varepsilon > 0$ and covers $\{q_\nu^i\}_{\nu \in \mathbb{N}}$ of S_i, $i \in \mathbb{N}$, with cubes q_ν^i, such that $\mu(\bigcup_{\nu \in \mathbb{N}} q_\nu^i) < \frac{\varepsilon}{2^i}, i \in \mathbb{N}$. Then the countable family $\{q_\nu^i\}_{i,\nu \in \mathbb{N}}$ covers $S := \bigcup_{i \in \mathbb{N}} S_i$ and satisfies

$$\mu\big(\cup_{i \in \mathbb{N}} S_i\big) \leq \mu\big(\cup_{i,\nu \in \mathbb{N}} q_\nu^i\big) < \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon \ .$$

□

EX. 2.1. *Hyperplanes $H^{n,k} = \{x \in \mathbb{R}^n \mid x_1 = \ldots = x_k = 0\}$ with $1 \leq k \leq n$, have Lebesgue measure zero.*

Proof. We only give the proof for $H^{2,1} = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0\}$. The extension of the proof to the general case is easy. Take the countable family of sets

$$S_i := \big\{(0, x_2) \mid i - 1 \leq x_2 \leq i\big\} \cup \big\{(0, x_2) \mid -i \leq x_2 \leq -(i-1)\big\} \ , \quad i \in \mathbb{N} \ .$$

The family $\{S_i\}_{i\in\mathbb{N}}$ covers the hyperplane $H^{2,1}$, so that by Lemma 2.1 the hyperplane has measure zero provided we prove that each set S_i has measure zero. To do so, fix i and chose $k \in \mathbb{N}$ arbitrarily, and put $\varepsilon = \frac{1}{k}$. With a partition $t_\nu = (i-1) + \frac{\nu}{k},\ \nu = 1, \ldots, k-1$, consider the cubes $q_\nu = \{(x_1, x_2) \mid |x_1| \leq \frac{1}{k},\ |x_2 - t_\nu| \leq \frac{1}{k}\}$. Then the family of cubes $q_\nu,\ \nu = 1, \ldots, k-1$, covers the set $\{(0, x_2) \mid i-1 \leq x_2 \leq i\}$ and we have

$$\mu(\{(0, x_2) \mid i-1 \leq x_2 \leq i\}) \leq \sum_{\nu=1}^{k-1} \mu(q_\nu) = \sum_{\nu=1}^{k-1} \frac{1}{k^2} < \frac{1}{k} = \varepsilon\ .$$

The same holds for the second set $\{(0, x_2) \mid -i \leq x_2 \leq -(i-1)\}$ of S_i. So S_i is of measure zero.

□

For continuous functions $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ it is possible to have $\mu(S) = 0$ and $\mu(f(S)) > 0$, even if $m \geq n$. Famous examples are the "Peano space-filling curves" given by continuous mappings $f : [0,1] \subset [0,1]^2 \to [0,1]^2$ which are surjective. Such a phenomenon is no more possible if f is a C^1-function.

THEOREM 2.1. *Let be given $S \subset U \subset \mathbb{R}^n$ with U open and $\mu(S) = 0$ (measure in $\mathbb{R}^n$). Suppose $f : U \to \mathbb{R}^m$ is a C^1-map with $m \geq n$. Then $\mu(f(S)) = 0$ (measure in $\mathbb{R}^m$).*

Proof. Firstly, we note that we can assume that U is covered by a countable family of (closed) cubes $q_i \subset U,\ i \in \mathbb{N}$, i.e., $U \subset \bigcup_{i\in\mathbb{N}} q_i$. Indeed we can, e.g., take the countable set of all points $r_i \in \operatorname{cl} U$ with rational coefficients and define q_i as the cubes with center points r_i and side length α for some $\alpha > 0$.
Since $\mu(S) = 0$ it follows that by putting $S_i := S \cap q_i$ we have $\mu(S_i) = 0$. In view of $f(S) \subset \cup_{i\in\mathbb{N}} f(S_i)$, by Lemma 2.1 it suffice to show, that $\mu(f(S_{i_0})) = 0$ for any fixed $i_0 \in \mathbb{N}$. Since $\mu(S_{i_0}) = 0$, by Definition 2.1, to any $\varepsilon > 0$, there is a cover of S_{i_0},

$$S_{i_0} \subset \bigcup_{j\in\mathbb{N}} q_{i_0}^j \qquad , \qquad \mu(\bigcup_{j\in\mathbb{N}} q_{i_0}^j) < \varepsilon\ ,$$

formed with cubes $q_{i_0}^j$ of measures $\mu(q_{i_0}^j) = \varepsilon_{i_0}^j,\ j \in \mathbb{N}$. We can assume (by considering the intersection $q_{i_0}^j \cap q_{i_0}$) , that $q_{i_0}^j \subset q_{i_0},\ j \in \mathbb{N}$. Obviously, we have proven the statement if we have shown that there is a constant $\gamma = \gamma_{i_0}$ such that for all $j \in \mathbb{N}$:

$$\mu(q_{i_0}^j) \leq \epsilon_{i_0}^j \qquad \Rightarrow \qquad \mu(f(q_{i_0}^j)) \leq \gamma \epsilon_{i_0}^j. \tag{2.1}$$

Take such a (nondegenerate) cube $\hat{q} := q_{i_0}^{j_0} \subset \mathbb{R}^n$ and put $\hat{\varepsilon} = \varepsilon_{i_0}^{j_0}$. Suppose $\hat{q}$ has side length $\hat{\alpha} > 0$. Since $\mu(\hat{q}) = \hat{\alpha}^n = \hat{\varepsilon}$, for the diameter $\hat{d} = \sqrt{n\hat{\alpha}^2}$ of $\hat{q}$ we find $\hat{d} = \sqrt{n}\sqrt[n]{\hat{\varepsilon}}$. Since $\hat{q} \subset q_{i_0}$ is convex, in view of the mean value formula in Lemma 7.1 it follows with $M := \max_{x\in q_{i_0}} \|\nabla f(x)\|$ that

$$\|f(x) - f(y)\| \leq M\|x - y\| \qquad \text{for all } x, y \in \hat{q}\ .$$

Thus, for all $x, y \in \hat{q}$,

$$\|f(x) - f(y)\| \leq M\hat{d} = M\sqrt{n}\sqrt[n]{\hat{\varepsilon}},$$

and $f(\hat{q})$ is contained in a ball $B \subset \mathbb{R}^m$ of radius $r := M\sqrt{n}\sqrt[n]{\hat{\varepsilon}}$ and volume $\frac{\pi^{m/2}}{\Gamma(\frac{m}{2}+1)}r^m$. Consequently, using $m \geq n$, and assuming $\hat{\varepsilon} \leq 1$ we find

$$\mu(f(\hat{q})) \leq \frac{\pi^{m/2}M^m n^{m/2}}{\Gamma(\frac{m}{2}+1)} \cdot \hat{\varepsilon}^{\frac{m}{n}} \leq \frac{\pi^{m/2}M^m n^{m/2}}{\Gamma(\frac{m}{2}+1)}\hat{\varepsilon} \ ,$$

and (2.1) is satisfied.

□

Note, that the statement of Theorem 2.1 needs no more be true in the case $m < n$. To provide a simple counterexample even of a C^∞-function $f : S \subset \mathbb{R}^n \to \mathbb{R}^m$ such that $\mu(S) = 0$ but $\mu(f(S)) > 0$, consider

$$f : S \subset \mathbb{R}^2 \to \mathbb{R} \ , \ f(x_1, x_2) = x_1 \ , \ S = \{x \in \mathbb{R}^2 \mid x_2 = 0\} \ \text{ with } f(S) = \mathbb{R} \ .$$

Recall the definition of generic and weakly generic sets in $\mathbb{R}^n$ (cf., Definition 1.1). From Lemma 2.1 we obtain

LEMMA 2.2. *Let $\{S_i\}_{i \in I}$, $I \subset \mathbb{N}$, be a (countable) family of weakly generic sets $S_i \subset \mathbb{R}^n$. Then the intersection*

$$S = \cap_{i \in I} S_i \quad \textit{is also weakly generic} \ .$$

Proof. By definition, the sets $C_i = \mathbb{R}^n \setminus S_i$ have measure zero. So by Lemma 2.1 also the set

$$\mathbb{R}^n \setminus S = \mathbb{R}^n \setminus (\cap_{i \in I} S_i) = \cup_{i \in I}(\mathbb{R}^n \setminus S_i) = \cup_{i \in I} C_i$$

has measure zero.

□

As for the openness, we only can allow a finite intersection of open sets. As a counterexample look at the infinite family of open sets $S_i = (-\frac{1}{i}, \frac{1}{i}), \ i \in \mathbb{N}$. The intersection is however closed:

$$\cap_{i \in \mathbb{N}} S_i = \{0\} \ .$$

EX. 2.2. *Let be given a finite family of open subsets $S_i \subset \mathbb{R}^n$. Then the (finite) intersection*

$$S = \cap_{i \in I} S_i \quad \textit{is also open in } \mathbb{R}^n.$$

As a corollary of Lemma 2.2 and Ex 2.2 we obtain a result that will be repeatedly used in the report.

COROLLARY 2.1. *Let be given a finite family of generic subsets $S_i \subset \mathbb{R}^n$, $i \in I, \ |I| < \infty$. Then also the (finite) intersection*

$$S = \cap_{i \in I} S_i \quad \textit{is generic in } \mathbb{R}^n \ .$$

For further definitions and basic properties of Lebesgue measure and Lebesgue integrable functions we refer to [34]. Here we only give *Fubini's formula* for (Lebesgue) integrable functions $f(x, y), x \in \mathbb{R}^n, y \in \mathbb{R}^k$:

$$\begin{aligned} \int_{A\times B} f(x,y)d\mu(\mathbb{R}^{n+k}) &= \int_A \left[\int_B f(x,y)d\mu(\mathbb{R}^k)\right] d\mu(\mathbb{R}^n) \\ &= \int_B \left[\int_A f(x,y)d\mu(\mathbb{R}^n)\right] d\mu(\mathbb{R}^k) \end{aligned} \tag{2.2}$$

where $A \subset \mathbb{R}^n, B \subset \mathbb{R}^k$ are Lebesgue measurable sets and $\mu(\mathbb{R}^k)$ denotes the measure on $\mathbb{R}^k$. We can apply this formula to prove

LEMMA 2.3. *Let $S \subset \mathbb{R}^k$ be generic in $\mathbb{R}^k$. Then the set*

$$\mathbb{R}^n \times S = \{(x,s) \mid x \in \mathbb{R}^n,\ s \in S\} \quad \textit{is generic in } \mathbb{R}^{n\times k} .$$

Proof. *weakly generic*: By assumption, $\mathbb{R}^k \setminus S$ has measure zero in $\mathbb{R}^k$. By formula (2.2) for the set $C = (\mathbb{R}^n \times \mathbb{R}^k) \setminus (\mathbb{R}^n \times S) = \mathbb{R}^n \times (\mathbb{R}^k \setminus S)$ we find with $f \equiv 1$,

$$\mu(C) = \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^k\setminus S} d\mu(\mathbb{R}^k)\right] d\mu(\mathbb{R}^n) = \int_{\mathbb{R}^n} [\mu(\mathbb{R}^k \setminus S)] d\mu(\mathbb{R}^n) = 0 .$$

openness: Take $(\overline{x}, \overline{s}) \in \mathbb{R}^n \times S$. By assumption, there exists some $\varepsilon > 0$ such that

$$s \in S \text{ holds for all } s \in \mathbb{R}^k \text{ with } \|s - \overline{s}\| < \varepsilon .$$

Now for all $(x, s) \in \mathbb{R}^n \times \mathbb{R}^k$ satisfying $\|(x, s) - (\overline{x}, \overline{s})\| < \varepsilon$, we find

$$\|s - \overline{s}\| \le \|(x,s) - (\overline{x}, \overline{s})\| < \varepsilon ,$$

i.e., $s \in S$, and thus $(x, s) \in \mathbb{R}^n \times S$.

□

EX. 2.3. *Let $z = (x, y), x \in \mathbb{R}^n, y \in \mathbb{R}^k$. Suppose that for any $x \in \mathbb{R}^n$ there is a weakly generic subset $S_x \subset \mathbb{R}^k$, i.e., $\mu(\mathbb{R}^k \setminus S_x) = 0$. Then the following set is weakly generic in $\mathbb{R}^{n+k}$:*

$$S_z := \cup_{x\in\mathbb{R}^n} (\{x\} \times S_x) .$$

Proof. By defining $C_x := \mathbb{R}^k \setminus S_x$ we have $C_z := \mathbb{R}^{n+k} \setminus S_z = \cup_{x\in\mathbb{R}^n} (\{x\} \times C_x)$. Now with the function $f(x,y) := \left\{\begin{smallmatrix} 1 & \text{if } y\in C_x \\ 0 & \text{otherwise}\end{smallmatrix}\right.$ we find from (2.2),

$$\mu(C_z) = \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^k} f(x,y) d\mu(\mathbb{R}^k)\right] d\mu(\mathbb{R}^n) = \int_{\mathbb{R}^n} [\mu(C_x)] d\mu(\mathbb{R}^n) = 0 .$$

So, $S_z \subset \mathbb{R}^{n+k}$ is a weakly generic subset of $\mathbb{R}^{n+k}$.

□

We emphasize that the same argument as in the preceeding proof cannot be used for the openness part. The fact that for any $x \in \mathbb{R}^n$ there is a generic (in particular open) subset $S_x \subset \mathbb{R}^k$, does not imply that the set $\cup_{x\in\mathbb{R}^n} (\{x\} \times S_x)$ is open in $\mathbb{R}^{n+k}$.

Open and dense vs generic.

A set $S \subset \mathbb{R}^n$ such that $\mu(\mathbb{R}^n \setminus S) = 0$, must be dense. Indeed, by definition of measure zero, such a set $\mathbb{R}^n \setminus S$ cannot contain any (nonempty) open set.
The converse is not true as can be seen from the example of the set $S = \mathbb{Q}$ which is dense in $\mathbb{R}$ with measure $\mu(\mathbb{Q}) = 0$ and $\mu(\mathbb{R} \setminus \mathbb{Q}) = \infty$. Interesting enough we even have:

$$S \subset \mathbb{R}^n \text{ is open and dense} \quad \not\Rightarrow \quad \mu(\mathbb{R}^n \setminus S) = 0 \; .$$

We provide an example: Take the (countable) set of rational numbers $\mathbb{Q}$ in $\mathbb{R}$,

$$\mathbb{Q} = \{q_1, q_2, q_3, \ldots\} = \{q_i \mid i \in \mathbb{N}\}$$

This set is dense in $\mathbb{R}$. Now choose $\varepsilon > 0$ (arbitrarily) and define, the open intervals $Q_i := (q_i - \varepsilon \frac{1}{2^i}, q_i + \varepsilon \frac{1}{2^i})$ of diameter $2\frac{1}{2^i}\varepsilon$ around q_i. Then the set

$$\mathcal{A} := \bigcup_{i \in \mathbb{N}} Q_i \qquad \text{is open and dense}$$

with $\mu(\mathcal{A}) \le \sum_{i \in \mathbb{N}} \mu(Q_i) \le 4\varepsilon$. Thus, this open, dense set $\mathcal{A}$ is not generic ($\mu(\mathbb{R} \setminus \mathcal{A}) = \infty$). Later in Section 2.3.2 we will prove that if S is a semi-algebraic set then density of S alone implies that $\operatorname{int} S$ is generic.

2.2. Generic sets in function spaces

In this section we introduce a special topology on the sets $C^k(\mathbb{R}^n, \mathbb{R}^m)$, $k \ge 1$. We often cannot go into details and we therefore refer, e.g., to [16, 18, 21] for further reading.

Let $X \subset \mathbb{R}^n$ be open. Often we will take $X = \mathbb{R}^n$. By $C^k(X, \mathbb{R})$, $k \ge 1$, we denote the space of all real-valued functions which are k-times continuously differentiable on X. $C^\infty(X, \mathbb{R})$ is the set of functions belonging to $C^k(X, \mathbb{R})$ for any $k \in \mathbb{N}$. As usual we identify

$$C^k(X, \mathbb{R}^m) \equiv [C^k(X, \mathbb{R})]^m \; .$$

To motivate our choice for a special topology on $C^k(X, \mathbb{R})$ ($k \ge 0$), let us look again at the function $f \in C^k(\mathbb{R}, \mathbb{R})$ given by

$$f(x) = \frac{1}{1 + x^2} \; \cos x \; ,$$

with infinitely many zeroes $\overline{x}_i, \; i \in \mathbb{Z}$ (see Figure 2.1). This function is nondegenerate in the sense that at any zero $\overline{x}_i$ the derivative satisfies

$$f'(\overline{x}_i) \neq 0 \qquad \forall i \in \mathbb{Z} \; .$$

Now, the topology on $C^k(X, \mathbb{R})$ should be such that there exists a neighborhoud U of f with the property: for all $g \in U$ the same nondegeneracy condition holds:

$$g'(\tilde{x}) \neq 0 \quad \text{for all } \tilde{x} \in \mathbb{R} \text{ satisfying } g(\tilde{x}) = 0 \; . \tag{2.3}$$

Let us for the moment restrain ourselfes to a compact subset $B = [-a, a]$, $a > 0$ fixed (see Figure 2.1). Seen as functions in $C^1(B, \mathbb{R})$ it is clear that we can chose some $\varepsilon > 0$, ε small enough, such that for all functions $g \in C^1(B, \mathbb{R})$ satisfying

$$|f(x) - g(x)| \; , \; |f'(x) - g'(x)| < \varepsilon \quad \forall x \in B \tag{2.4}$$

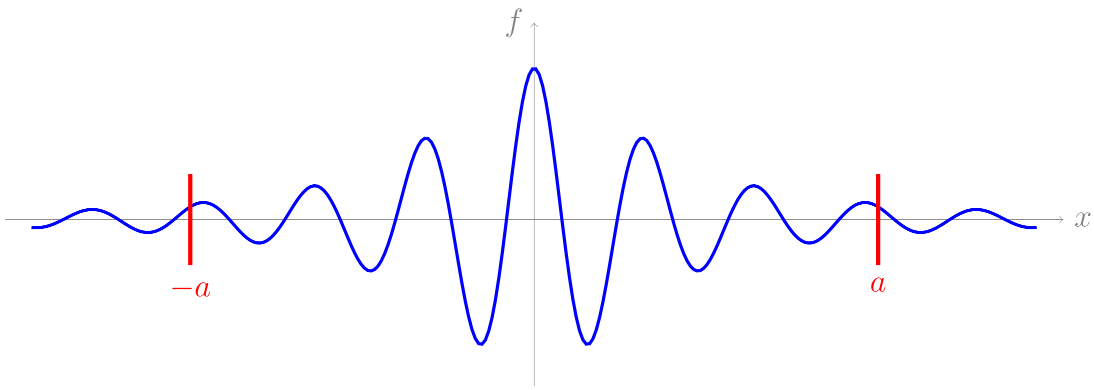

FIGURE 2.1. Function $f(x) = \frac{1}{x^2+1} \ \cos x$.

it holds,

$$g'(\tilde{x}) \neq 0 \quad \text{for all } \tilde{x} \in B \text{ satisfying } g(\tilde{x}) = 0 \ .$$

However this is no more true if we take $B = \mathbb{R}$. Obviously for any constant $\varepsilon > 0$ there

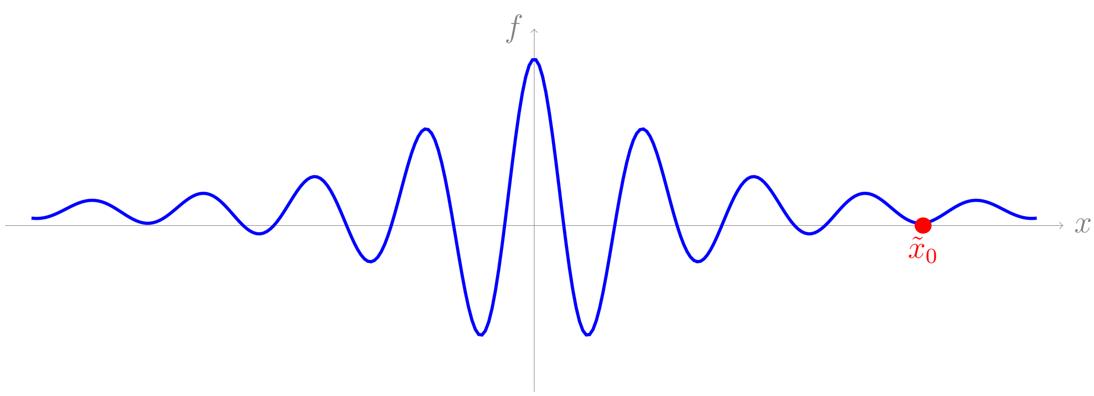

FIGURE 2.2. function $f(x) = \frac{1}{x^2+1} \ \cos x + \varepsilon_0$

exists some ε_0, $0 < \varepsilon_0 < \varepsilon$, such that the function $\tilde{g} = f + \varepsilon_0$ has the form as indicated in Figure 2.2, with a point $\tilde{x}_0$ satisfying $\tilde{g}(\tilde{x}_0) = \tilde{g}'(\tilde{x}_0) = 0$, and (2.3) does not hold at $\tilde{x}_0$. So the ε neigborhood of f as defined in (2.4) does not capture the behaviour of "$g \approx f$ near ∞".

To circumvent this problem, instead of defining an ε-neigborhood of f by a constant $\varepsilon > 0$ (as in (2.4)), we define a ϕ-neighborhood of f by a function $\phi : \mathbb{R} \to \mathbb{R}$, $\phi \in C^0(\mathbb{R}, \mathbb{R})$, $\phi(x) > 0 \ \forall x \in \mathbb{R}$. Then by chosing, e.g., $\phi(x) = \frac{1}{2} \frac{1}{1+x^2}$, and a neigborhood of f via

$$|f(x) - g(x)| \ , \ \ |f'(x) - g'(x)| < \frac{1}{2} \frac{1}{1+x^2} \qquad \forall x \in \mathbb{R}^n \ , \tag{2.5}$$

then for all functions $g \in C^1(\mathbb{R}, \mathbb{R})$ satisfying this relation also the nondegeneracy condition (2.3) will hold.

In this report instead of considering functions on compact sets we prefer to develop our theory for functions $C^k(\mathbb{R}^n, \mathbb{R})$ defined on the whole $\mathbb{R}^n$. To this end we introduce the so-called strong topology on $C^k(\mathbb{R}^n, \mathbb{R})$ (cf., also [21, Def.6.2.8]). Recall that for $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, $|\alpha| = \sum_{i=1}^n \alpha_i$, the expression $\partial^\alpha f(x)$ denotes the α-partial derivative

$$\partial^\alpha f(x) = \frac{\partial^{|\alpha|}}{\partial_{x_n}^{\alpha_n} \ldots \partial_{x_1}^{\alpha_1}} f(x) \ .$$

Definition 2.2. *For fixed $k \in \mathbb{N}$, a basis of open sets (neigborhoods) for the (strong) C^k_s-topology on $C^k(\mathbb{R}^n, \mathbb{R})$ is given by all sets $U^k_{\phi,f}$,*

$$U^k_{\phi,f} = \{g \in C^k(\mathbb{R}^n, \mathbb{R}) \mid |\partial^\alpha f(x) - \partial^\alpha g(x)| < \phi(x), \ \forall x \in \mathbb{R}^n \ , \ \forall \alpha, \ |\alpha| \le k\} \ ,$$

where ϕ is a function in $C^0_+(\mathbb{R}^n, \mathbb{R}) := \{\phi \in C^0(\mathbb{R}^n, \mathbb{R}) \mid \phi(x) > 0 \ \forall x \in \mathbb{R}^n\}$ and $f \in C^k(\mathbb{R}^n, \mathbb{R})$.
In an obvious way, for $\ell \ge 0, \ \ell \le k \le \infty$, a C^ℓ_s-topology can be defined on $C^k(\mathbb{R}^n, \mathbb{R})$. A C^∞_s-topology on $C^\infty(\mathbb{R}^n, \mathbb{R})$ is defined by all neigborhoods $U^k_{\phi,f}$, $k \in \mathbb{N}, \phi \in C^0_+(\mathbb{R}^n, \mathbb{R})$, $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$.
The C^k_s-topology on $[C^k(\mathbb{R}^n, \mathbb{R})]^m$ is defined as the product topology. More precisely, for $F = (f_1, \ldots, f_m) \in [C^k(\mathbb{R}^n, \mathbb{R})]^m$, the neigborhoods $U^k_{\phi,F}$ are defined by (see Ex. 2.4)

$$U^k_{\phi,F} := \otimes_{j=1}^m U^k_{\phi,f_j} \ .$$

Ex. 2.4. *Let $F = (f_1, \ldots, f_m) \in [C^k(\mathbb{R}^n, \mathbb{R})]^m$. Show that for given $\phi_j \in C^0_+(\mathbb{R}^n, \mathbb{R})$, $j = 1, \ldots, m$, we have*

$$U^k_{\phi,F} := \otimes_{j=1}^m U^k_{\phi,f_j} \subset \otimes_{j=1}^m U^k_{\phi_j,f_j} \ ,$$

if we define

$$\phi(x) := \min_{1 \le j \le m} \phi_j(x) \in C^0_+(\mathbb{R}^n, \mathbb{R}) \ .$$

Remark 2.1. We notice that if a subset S of $C^\infty(\mathbb{R}^n, \mathbb{R})$ is dense in $C^\infty(\mathbb{R}^n, \mathbb{R})$ wrt. the C^k_s-topology ($k \in \mathbb{N}$) then S is dense in $C^\infty(\mathbb{R}^n, \mathbb{R})$ wrt. the C^ℓ_s-topology for all $\ell \in \mathbb{N}$, $\ell \le k$.
Indeed, density of S wrt. the C^k_s-topology means that given any $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$, for any neigborhood $U^k_{\phi,f}$ (i.e., for any $\phi \in C^0_+(\mathbb{R}^n, \mathbb{R})$) there exists some $f_\phi \in S$ such that

$$|\partial^\alpha f(x) - \partial^\alpha f_\phi(x)| < \phi(x) \qquad \forall x \in \mathbb{R}^n, \ \forall \alpha, \ |\alpha| \le k \ .$$

Then clearly also for all $\ell \le k$ it follows,

$$|\partial^\alpha f(x) - \partial^\alpha f_\phi(x)| < \phi(x) \qquad \forall x \in \mathbb{R}^n, \ \forall \alpha, \ |\alpha| \le \ell \ ,$$

i.e., $f_\phi \in U^\ell_{\phi,f}$. Note that such a statement is not true for openness.

Remark 2.2. *A warning:* The spaces $C^k(\mathbb{R}^n, \mathbb{R})$ endowed with the C^k_s-topology are not topological vector spaces. By definition, in a topological vector space the mappings

$$\begin{array}{rl} C^k(\mathbb{R}^n, \mathbb{R}) \times C^k(\mathbb{R}^n, \mathbb{R}) \to C^k(\mathbb{R}^n, \mathbb{R}) & \quad (f, g) \to f + g \\ \mathbb{R} \times C^k(\mathbb{R}^n, \mathbb{R}) \to C^k(\mathbb{R}^n, \mathbb{R}) & \quad (\lambda, f) \to \lambda f \end{array}$$

should be continuous wrt. the C^k_s-topology (see,e.g., [21, Rem.6.1.10]). The first mapping is continuous (proof as an exercise) but the second is not. Take, e.g., $f \equiv 1$, $\overline{\lambda} = 0$, $\overline{f} := \overline{\lambda} \cdot 1 = 0$, the sequences $\lambda_n = \frac{1}{n}$, $f_n := \lambda_n \cdot f$, and the function $\phi(x) = e^{-x^2}$. Then $\lambda_n f$ is not continuous in the neigborhood $U^k_{\phi,\overline{f}}$ of $\overline{f}$. In other words, wrt. C^k_s,

$$f_n \nrightarrow \overline{f} = 0 \ .$$

It can be shown (see [21, p.306]) that for all $k \in \mathbb{N} \cup \{\infty\}$ the spaces

$$C^k(\mathbb{R}^n, \mathbb{R}) \ \text{ endowed with the } C^k_s\text{-topology are Baire spaces} \ ,$$

i.e., they have the following property: If $\mathcal{A}_i \subset C^k(\mathbb{R}^n, \mathbb{R})$, $i \in \mathbb{N}$, are C^k_s-open and C^k_s-dense then

$$\mathcal{A} = \cap_{i\in\mathbb{N}} \mathcal{A}_i \quad \text{are } C^k_s\text{-dense} \ .$$

We now can define what we mean by genericity in function spaces (see, e.g., [14, p.127], [21, p.306]).

DEFINITION 2.3. [generic set in $C^k(\mathbb{R}^n, \mathbb{R})$]
Let X be a Baire space (such as $C^k(\mathbb{R}^n, \mathbb{R})$ with the C^k_s-topology). Then a subset $\mathcal{A} \subset X$ is called generic in X (or residual in X, cf., [18, p.74]) *if $\mathcal{A}$ contains a countable intersection of open and dense sets $\mathcal{A}_i \subset X$. In particular a generic set is dense but need not be open. Note that by definition, in a Baire space, a countable intersection of generic sets is still a generic set.*
We emphasize that in particular, a subset $\mathcal{A} \subset C^k(\mathbb{R}^n, \mathbb{R})$ (endowed with the C^k_s-topology) is generic in $C^k(\mathbb{R}^n, \mathbb{R})$ if $\mathcal{A}$ is open and dense in $C^k(\mathbb{R}^n, \mathbb{R})$ wrt. the C^k_s-topology.

Lateron, we regularly will make use of the following norm $\| \cdot \|_{k,C}$ for $f \in C^k(\mathbb{R}^n, \mathbb{R})$, resp., $F = (f_1, \dots, f_m) \in [C^k(\mathbb{R}^n, \mathbb{R})]^m$, $k \in \mathbb{N}$, and compact $C \subset \mathbb{R}^n$:

$$\|f\|_{k,C} := \max_{|\alpha|\le k} \ \max_{x\in C} \ |\partial^\alpha f(x)| \quad , \quad \|F\|_{k,C} := \max_{0\le j\le m} \ \|f_j\|_{k,C} \ . \tag{2.6}$$

2.3. Algebraic and semi-algebraic sets

This section presents definitions and basic properties of algebraic and semi-algebraic sets. Algebraic sets are solution sets of polynomial equations, and semi-algebraic sets are defined by polynomial equalities and inequalities.
A polynomial function on $\mathbb{R}^n$ of degree $\le k$ is a function $p : \mathbb{R}^n \to \mathbb{R}$ of the form

$$p(x) = \sum_{|\alpha|\le k} c_\alpha x^\alpha \ ,$$

where the sum is taken over all vectors $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, with $|\alpha| := \sum_i \alpha_i \le k$, and x^α stands for $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$. We write $p \ne 0$ if at least one of the coefficients c_α is not equal to zero or equivalently if $p(\overline{x}) \ne 0$ holds for some $\overline{x} \in \mathbb{R}^n$.

EX. 2.5. *Show that for polynomials we have:*

$$p(x) = 0 \ \ \forall x \text{ in a nonempty open subset } U \text{ of } \mathbb{R}^n \quad \Rightarrow \quad p(x) = 0 \ \ \forall x \in \mathbb{R}^n.$$

Proof. Let the open set $U \subset \mathbb{R}^n$ contain a point $\overline{x}$. By Taylor expansion, p has the respresentation ($\alpha! = \alpha_1! \cdots \alpha_n!$)

$$p(x) = \sum_{|\alpha| \le k} \frac{\partial^\alpha p(\overline{x})}{\alpha!} (x - \overline{x})^\alpha \ .$$

In view of $p(x) = 0$, $x \in U$, it follows $\frac{\partial}{\partial x_j} p(\overline{x}) = \lim_{h \to 0} \frac{p(\overline{x} + h e_j) - p(\overline{x})}{h} = 0$, $j = 1, \ldots, n$, and further in the same way $\partial^{(\alpha)} p(\overline{x}) = 0$ for all $|\alpha| \le k$. So we find $p \equiv 0$.

□

DEFINITION 2.4. *A set $S \subset \mathbb{R}^n$ is called algebraic if it can be written as*

$$S = \{x \in \mathbb{R}^n \mid p_1(x) = 0, \ldots, p_k(x) = 0\}, \tag{2.7}$$

with a finite number of polynomial functions $p_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, k$.
A set $S \subset \mathbb{R}^n$ is called semi-algebraic if it can be written as a finite union of sets of the following form:

$$V = \{x \in \mathbb{R}^n \mid p_1(x) = 0, \ldots, p_k(x) = 0; \ \ q_1(x) > 0, \ldots, q_m(x) > 0\}, \tag{2.8}$$

with a finite number of polynomial functions $p_j, q_\ell : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, k$; $\ell = 1, \ldots, m$.

We present two examples. Denote by M^n the set of all real $(n \times n)$-matrices, $M^n \equiv \mathbb{R}^{n \times n}$.

- The set
$$S_1 = \{A \in M^n \mid \det(A) = 0\}$$
is algebraic, as $p(A) = \det(A)$ (cf., (7.1)) is a polynomial function.
- The complement $S_2 = S_1^c$,
$$S_2 = \{A \in M^n \mid \det(A) \neq 0\} \ ,$$
is a semi-algebraic set, as S_2 can be written as union of semi-algebraic sets, $S_2 = \{A \in M^n \mid \det(A) > 0\} \cup \{A \in M^n \mid -\det(A) > 0\}$.

2.3.1. Properties of semi-algebraic sets.

EX. 2.6. *The sets V in Definition 2.4 of semi-algebraic sets may also contain finitely many inequality constraints $h_1(x) \ge 0, \ldots, h_s(x) \ge 0$.*

Hint: Use $\{x \mid h(x) \ge 0\} = \{x \mid h(x) = 0\} \cup \{x \mid h(x) > 0\}$.

EX. 2.7. *Let $S_1, S_2 \subset \mathbb{R}^n$ be semi-algebraic. Then the intersection $S_1 \cap S_2$ is semi-algebraic.*

Proof. Let $S_i = \cup_{\nu_i \in J_i} V^i_{\nu_i}$, with finite index sets J_i, $i = 1, 2$ ($V^i_{\nu_i}$ of the form (2.8)). Obviously for each pair $\nu_1 \in J_1,\ \nu_2 \in J_2$, the set $V^1_{\nu_1} \cap V^2_{\nu_2}$ is semi-algebraic (take the $=, >$'s of both sets $V^1_{\nu_1}, V^2_{\nu_2}$). The statement then follows by the formula

$$S_1 \cap S_1 = \left(\cup_{\nu_1 \in J_1} V^1_{\nu_1}\right) \bigcap \left(\cup_{\nu_2 \in J_2} V^2_{\nu_2}\right) = \cup_{\substack{\nu_1 \in J_1 \\ \nu_2 \in J_2}} \left(V^1_{\nu_1} \cap V^2_{\nu_2}\right) .$$

□

By definition, for a polynomial p the set $p^{-1}(0)$ is algebraic. More generally we have

EX. 2.8. *Let $S \subset \mathbb{R}^n$ be a semi-algebraic set, and let $h : \mathbb{R}^p \to \mathbb{R}^n$ be a polynomial function. Then the inverse image $h^{-1}(S) = \{y \in \mathbb{R}^p \mid h(y) = x, x \in S\}$ is a semi-algebraic subset of $\mathbb{R}^p$.*

Hint: E.g., for the case $S = \{x \in \mathbb{R}^n \mid p(x) = 0\}$ we find $h^{-1}(S) = \{y \in \mathbb{R}^p \mid h(y) = x, x \in S\} = \{y \in \mathbb{R}^p \mid (p \circ h)(y) = 0\}$.

REMARK 2.3. By definition, algebraic sets are closed, whereas semi-algebraic sets may be neither closed nor open.

Recall the example $\mathbb{Q} \subset \mathbb{R}$ which shows that density of a set $S \subset \mathbb{R}$ does not imply weak genericity (nor openness). So a dense subset $S \subset \mathbb{R}^n$ can be "far" from being generic. However if S is semi-algebraic then density of S alone implies genericity of $\operatorname{int} S$.

THEOREM 2.2. *Let $S \subset \mathbb{R}^n$ be a semi-algebraic set. If S is dense in $\mathbb{R}^n$ then $\mathbb{R}^n \setminus S$ has Lebesgue measure zero in $\mathbb{R}^n$. Moreover, the set $\operatorname{int} S$ is a generic subset of $\mathbb{R}^n$.*

Proof. We provide an elementary proof due to David Preinerstorfer (private communication) which is only based on Theorem 3.1.

By definition, the semi-algebraic set is given by $S = \bigcup_{i=1}^{\ell} S_i$, with

$$S_i = \{x \in \mathbb{R}^n \mid p_{ij}(x) = 0,\ j = 1, \dots, k_i\ ;\ q_{ij}(x) > 0,\ j = 1, \dots, m_i\}, \quad i = 1, \dots, \ell\ ,$$

where k_i and m_i are positive integers, p_{ij} and q_{ij} are polynomials, and $S_i \neq \emptyset$ for all i. We next show that

$$\text{bd}\,(S) \subseteq \bigcup_{i=1}^{\ell} \text{bd}\,(S_i) \qquad \text{has Lebesgue measure zero.}$$

For i such that $p_{ij} \not\equiv 0$ for at least one j we have

$$\text{bd}\,(S_i) \subseteq \text{cl}\,(S_i) \subseteq p_{ij}^{-1}(0)\ ,$$

which by Theorem 3.1 satisfies $\mu(\text{bd}\,(S_i)) = 0$. For all other i, we have

$$S_i = \{x \in \mathbb{R}^n \mid q_{ij}(x) > 0\ \forall j = 1, \dots, m_i\}\ ,$$

with $q_{ij} \not\equiv 0$ for all j as $S_i \neq \emptyset$. One can easily show that

$$\text{bd}\,(S_i) \subseteq \bigcup_{j=1}^{m_i} q_{ij}^{-1}(0)\,,$$

and therefore also in this case $\mu(\text{bd}\,(S_i)) = 0$. Thus we have $\mu(\text{bd}\,(S)) = 0$. Moreover for any dense set $S \subset \mathbb{R}^n$ the complement of int (S) coincides with bd (S), *i.e.*,

$$S \text{ dense} \quad \Rightarrow \quad (\mathbb{R}^n \setminus \text{int}\,(S)) = \text{bd}\,(S)\;. \tag{2.9}$$

This proves that the set int (S) is a generic subset of $\mathbb{R}^n$. To prove (2.9) we only note that for dense S we have cl $(S) = \mathbb{R}^n$ so that

$$\text{bd}\,(S) := \text{cl}\,(S) \setminus \text{int}\,(S) = \mathbb{R}^n \setminus \text{int}\,(S)\;.$$

□

We also present some further deeper results on semi-algebraic sets.

Other properties of semi-algebraic sets. Let $S \subset \mathbb{R}^n$ be semi-algebraic.

(1) Then the sets cl S, int S, bd S are semi-algebraic.

(2) S has only finitely many connected components, each of which is semi-algebraic.

(3) Let A be a semi-algebraic subset of vectors $(x, y) \in \mathbb{R}^n \times \mathbb{R}^k$. Then the projection proj$_x A = \{x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^k$ such that $(x, y) \in A\}$ of A onto $\mathbb{R}^n$ is a semi-algebraic set in $\mathbb{R}^n$.

(4) *[Tarski-Seidenberg]* Let $p : \mathbb{R}^n \to \mathbb{R}^p$ be a polynomial mapping. Then the image $p(S)$ is a semi-algebraic set.

For proofs see, e.g., [14, p.18] for (1),(2) and [10, Cor. 2.4] for (3),(4).

Later in Theorem 5.9, we will see that any semi-algebraic set allows a complete (stratification) partition into smooth manifolds.
We add two illustrative examples. Recall that a set $C \subset \mathbb{R}^n$ is a cone if $x \in C$ implies $\lambda x \in C$ for all $0 \leq \lambda \in \mathbb{R}$. Let S^n denote the set of real symmetric $(n \times n)$-matrices, $S^n \equiv \mathbb{R}^{(n+1)n/2}$.

Ex. 2.9. *The set*

$$\mathcal{CP}^n = \{A \in S^n \mid A = \sum_{i=1}^{p} b_i b_i^T,\ 0 \leq b_i \in \mathbb{R}^n,\ i = 1, \ldots, p,\ \textit{for some } p\}$$

is called the cone of completely positive matrices. In the definition of $\mathcal{CP}^n$, wlog., we can assume $p \leq \frac{(n+1)n}{2}$. This follows by Caratheodory's Theorem (see, [13, Theorem 3.6]) *which implies that each element A in the cone $\mathcal{CP}^n \subset \mathbb{R}^{(n+1)n/2}$ can be written as a conic combination $A = \sum_{i=1}^p \lambda_i b_i b_i^T$, $\lambda_i \geq 0$, $b_i \geq 0$, of at most $p \leq \frac{(n+1)n}{2}$ linearly independent elements $b_i b_i^T$.*

Show that $\mathcal{CP}^n$ is a semi-algebraic set in S^n.

Proof. The sets

$$V_p = \{(A, b_1, \ldots, b_p) \in S^n \times \mathbb{R}^{n\times p} \mid A = \sum_{i=1}^{p} b_i b_i^T,\ 0 \le b_i,\ i = 1, \ldots, p\}\ ,$$

$1 \le p \le \frac{(n+1)n}{2}$, are semi-algebraic as solution sets of polynomial $=, \ge$'s. By property (3) above also the projections proj $_A V_p$,

$$\text{proj}\,_A V_p = \{A \in S^n \mid \exists b_i \in \mathbb{R}^n \text{ such that } (A, b_1, \ldots, b_p) \in V_p\},\ 1 \le p \le \frac{(n+1)n}{2},$$

are semi-algebraic. Since $\mathcal{CP}^n$ is the union of these sets proj $_A V_p$, also $\mathcal{CP}^n$ is semi-algebraic.

□

In [5] by applying Theorem 2.2, genericity results on the so-called cp-rank of matrices in $\mathcal{CP}^n$ are obtained.

Ex. 2.10. *Let $C \subset \mathbb{R}^n$ be a semi-algebraic cone. Then the dual cone C^* of C,*

$$C^* = \{y \in \mathbb{R}^n \mid c^T y \ge 0 \ \ \forall c \in C\}\ ,$$

is also semi-algebraic.

Proof. Consider the sets

$$S = \{y \in \mathbb{R}^n \mid \exists c \in C \ \text{ with } \ c^T y < 0\}$$

and

$$S_1 = \{(y, c) \in \mathbb{R}^n \times \mathbb{R}^n \mid c^T y < 0\}\ , \qquad S_2 = \{(y, c) \in \mathbb{R}^n \times \mathbb{R}^n \mid c \in C\}\ .$$

The sets S_1, S_2 are semi-algebraic and by Ex. 2.7 also the set

$$S_0 := S_1 \cap S_2 = \{(y, c) \in \mathbb{R}^n \times \mathbb{R}^n \mid c \in C,\ c^T y < 0\}$$

is semi-algebraic. In view of property (3) above, also the projection $S =$ proj $_y S_0$ is semi-algebraic. Since C^* is the complement S^c of S, by property (1) above also C^* must be semi-algebraic.

□

2.3.2. Properties of algebraic sets.

In this subsection we consider algebraic sets, *i.e.*, semi-algebraic sets that are defined only by equalities (see Definition 2.4). We denote by $R[x]$ the set (ring) of all polynomials in the variable $x \in \mathbb{R}^n$.

An algebraic set $V = \{x \in \mathbb{R}^n \mid p_1(x) = 0, \ldots, p_k(x) = 0\}$ will often be denoted by $V(p_1, \ldots, p_k)$.

We are interested in the geometric structure of such sets. This is the topic of the field of Algebraic Geometry. The structure of solution sets of (general) C^1-functions is investigated in Differential Geometry.

A first observation explains the difference between Algebraic- and Differential-Geometry. It is easy to construct C^∞-functions $f \neq 0$, such that the solution set $S = \{x \in \mathbb{R}^n \mid f(x) = 0\}$ contains (nonempty) open sets. See Theorem 5.3 in Section 5.1. According to Ex 2.5 this is not possible for polynomial functions.

We now give some illustrative examples of algebraic sets. They in particular show that in general an algebraic set needs not be a manifold (see Definition 5.1). However these sets allow a partition into manifolds (see Theorem 5.9).

Ex. 2.11. $V = V(x_2^2 - x_1^2(x_1 - 1))$ *(see Figure 2.3).*

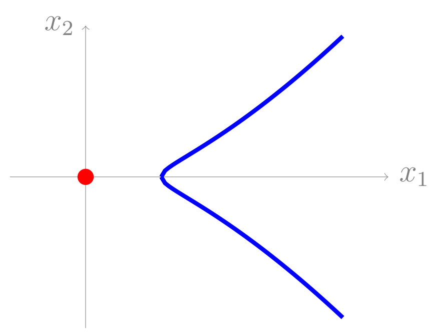

FIGURE 2.3. Algebraic set defined by $x_2^2 - x_1^2(x_1 - 1) = 0$ with an isolated point at $(0, 0)$.

Ex. 2.12. $V = V(p_1, p_2)$ *with* $p_1(x) = x_1x_2,\ p_2(x) = x_1x_3$.
The gradients of p_1, p_2 *are* $\nabla p_1(x) = (x_2, x_1, 0), \nabla p_2(x) = (x_3, 0, x_1)$. *At the point* $\overline{x} = (1, 0, 0) \in V$ *the gradients*

$$\nabla p_1(\overline{x}) = (0, 1, 0), \nabla p_2(\overline{x}) = (0, 0, 1) \quad \textit{are linearly independent}.$$

So (by the Implicit Function Theorem 7.2), locally around $\overline{x}$*, the set* V *consists of a* 1*-dimensional manifold given by* $\{x \in \mathbb{R}^3 \mid x_2 = x_3 = 0\}$*. However, at the point* $\tilde{x} = (0, 1, 1) \in V$*, the gradients* $\nabla p_1(\tilde{x}) = (1, 0, 0), \nabla p_2(\tilde{x}) = (1, 0, 0)$ *are linearly dependent. Around* $\tilde{x}$ *the set* V *is a* 2*-dimensional manifold given by* $\{x \in \mathbb{R}^3 \mid x_1 = 0\}$.

Ex 2.12 shows that for algebraic sets $V(p_1, \ldots, p_k)$ the following (bad situation) is possible:

- There are algebraic sets $V = V(p_1, \ldots, p_k) \subset \mathbb{R}^n$ which locally around $\overline{x} \in V$ define a manifold of dimenion $n - k$ but V contains other parts of higher dimension. See, e.g., Ex.2.12.

Ex. 2.13. $V = V(p_1, p_2)$ *with* $p_1(x) = x_1^2 - x_2x_3,\ p_2(x) = x_1(x_3 - 1)$.

Preciser information about the structure of an algebraic set can be obtained by techniques from Algebra. We only give some more details connected with the notion of reducibility of an algebraic set or the concept of ideals. For further reading we refer to books on Algebraic Geometry (see *e.g.*, [3]).
Let us emphasize that the number of equations defining V, alone, does not give any (non-trivial) information about the algebraic set V. Indeed the set $V = V(p_1, \dots, p_k)$ can be written equivalently as

$$V = \{x \in \mathbb{R}^n \mid p(x) := \sum_{i=1}^{k} p_i^2(x) = 0\}.$$

DEFINITION 2.5. *An algebraic set $V \subset \mathbb{R}^n$ is called reducible if V can be decomposed as $V = V_1 \cup V_2$ with algebraic sets V_1, V_2 satisfying $V_1 \neq V, V_2 \neq V$.*
An algebraic set V which is not reducible is called irreducible.

The set $V(x_1x_2, x_1x_3)$ of Ex 2.12 is reducible as it can be decomposed as

$$V = V_1 \cup V_2 \quad \text{with} \quad V_1 = \{x \mid x_1 = 0\}, \;\; V_2 = \{x \mid x_2 = x_3 = 0\} \; .$$

Note however that this decomposition does not define a partition (see Definition 5.3), since $V_1 \cap V_2 = \{0\}$ is nonempty. A partition of V into manifolds is, *e.g.*, given by

$$V = \{x \mid x_1 = 0\} \cup \{x \mid x_2 = x_3 = 0, x_1 > 0\} \cup \{x \mid x_2 = x_3 = 0, x_1 < 0\} \; .$$

However the last 2 sets are not algebraic but only semi-algebraic. Many other partitions are possible.

EX. 2.14. *Give a decomposition into irreducible algebraic subsets of the algebraic set V in Example 2.13.*

Hint: Show $V = \{x \mid x_1 = 0, x_2 \cdot x_3 = 0\} \cup \{x \mid x_3 = 1, \; x_2 = x_1^2\}$.

One can show that the algebraic set V of Example 2.11 is irreducible.

As a consequence of the famous Hilbert Basis Theorem one can obtain the following decomposition result (see *e.g.*, [3, 3.1.5 Proposition]).

PROPOSITION 2.1. *p-13-1 Any algebraic set $V \subset \mathbb{R}^n$ can be decomposed as*

$$V = V_1 \cup \ldots \cup V_s \quad \textit{with irrecucible sets } V_i.$$

If we remove the sets V_i that are contained in another set V_j, the decomposition obtained is unique.

Recall that the sets V_i of this decomposition need not to be manifolds.

Since the same algebraic set can be defined by different sets of polynomials p_i the question arises whether there is some sort of "basic set". To answer this, we have to consider the ideal $I(V)$ of an algebraic set:

$$I(V) := \{p \in R[x] \mid p(x) = 0 \;\text{ holds for every } x \in V\;\}. \tag{2.10}$$

DEFINITION 2.6. *A subset I of a ring R is called an ideal if:*

(1) $0 \in I$
(2) $p, q \in I \quad \Rightarrow \quad p + q \in I.$
(3) $p \in I, q \in R \quad \Rightarrow \quad p \cdot q \in I.$

Obviously the set $I(V) \subset R[x]$ in (2.10) is an ideal in $R[x]$ (proof as Exercise).

Consider now with explicitly given polynomials p_i the algebraic set $V(p_1, \ldots, p_k)$. We can define the ideal

$$I(p_1, \ldots, p_k) = \left\{ \sum_{i=1}^{k} p_i q_i \mid q_i \in R[x] \right\} .$$

EX. 2.15. *Show that $I(p_1, \ldots, p_k)$ is an ideal in $R[x]$ satisfying*

$$I(p_1, \ldots, p_k) \subset I(V(p_1, \ldots, p_k)) .$$

In general the equality needs not to hold for the ideals in Exercise 2.15. Take for example $V(p_1) \subset \mathbb{R}^2$ with $p_1(x) = x_1^2 + x_2^2$. Then $V(p_1) = \{0\}$ and $p = x_1$ (or $p = x_2$) is in $I(V(p_1))$ but obviously x_1 (or x_2) are not in $I(x_1^2 + x_2^2)$.

However with the help of the Hilbert Basis Theorem one can prove that to each algebraic set V there is a set of polynomials generating $I(V)$ (see *e.g.*, [3, Section 3.1]).

PROPOSITION 2.2. *p-13-2 Let $V \subset \mathbb{R}^n$ be an algebraic set. Then there exist polynomials $P_1, \ldots, P_s$ such that $I(V) = I(P_1, \ldots, P_s)$ and thus $V = V(P_1, \ldots, P_s)$.*

EX. 2.16. *Show that for an algebraic set V the relation $I(V) = I(P_1 \ldots, P_s)$ implies $V = V(P_1, \ldots, P_s)$.*

Proof. Let $V = V(p_1, \ldots, p_k)$. Since $p_i \in I(V) = I(P_1 \ldots, P_s)$, each polynomial p_i has a representation $p_i = \sum_{j=1}^{s} P_j q_{ji}$. So $x \in V(P_1, \ldots, P_s)$ implies $x \in V(p_1, \ldots, p_k)$. On the other hand we have $P_j \in I(V)$ and thus $x \in V$ implies $P_j(x) = 0$. So we have proven $V = V(p_1, \ldots, p_k) = V(P_1, \ldots, P_s)$ □

One can show for example, that for the set $V = V(x_1^2 + x_2^2)$ we have $I(V(x_1^2 + x_2^2)) = I(x_1, x_2)$.

Also the reducibility of an algebraic set V is connected with the set $I(V)$ as follows.

DEFINITION 2.7. *Let I be an ideal in a commutative ring R. Then I is called a prime ideal if*

(1) $I \neq R$
(2) $p, q \in R$, $p \cdot q \in I \;\Rightarrow\; p \in I$ *or* $q \in I$.

Take, *e.g.*, $V = V(x_1 \cdot x_2)$ and consider its ideal

$$I(V(x_1 \cdot x_2)) = \{p \in R[x] \mid p(x) = 0 \; \forall x \in V(x_1 \cdot x_2)\} .$$

Then for the polynomials x_1, x_2 we have $p := x_1 \cdot x_2 \in I(V(x_1 \cdot x_2))$ but $x_1, x_2 \notin I(V(x_1 \cdot x_2))$. So the ideal $I(V(x_1 \cdot x_2))$ is not prime. Note that V is reducible, $V = \{x \mid x_1 = 0\} \cup \{x \mid x_2 = 0\}$ The following characterisation holds.

PROPOSITION 2.3. *p-13-3 The algebraic set $V \subset \mathbb{R}^n$ is irredicible if and only if the ideal $I(V)$ is prime.*

We end up this subsection with some observations.

- Recall (example after Ex. 2.15) that in general $I(p_1, \ldots, p_k)$ is not equal to $I(V(p_1, \ldots, p_k))$. Also, even if $I(p_1, \ldots, p_k)$ is prime, the ideal $I(V(p_1, \ldots, p_k))$ need not to be prime (i.e., $V(p_1, \ldots, p_k)$ need not to be irreducible). An example is given by $p = x_1^2(x_1 - 1)^2 + x_2^2$. Here one can show that $I(p)$ is prime but $I(V(p))$ not, and thus $V(p) = \{(0,0), (1,0)\}$ is reducible.

- One can show that the set $V(x_1^2(x_1 - 1) + x_2^2)$ is irreducible.

CHAPTER 3

Genericity results in linear algebra

3.1. Basic results

In this section we present first genericity results in linear algebra. They will be based on the following statement on the set of zeroes of polynomial functions.

THEOREM 3.1. *Let $p : \mathbb{R}^N \to \mathbb{R}$ be a polynomial mapping, $p \neq 0$. Then the set $p^{-1}(0) = \{x \in \mathbb{R}^N \mid p(x) = 0\}$ has (Lebesgue) measure zero in $\mathbb{R}^N$. Moreover, since $p^{-1}(0)$ is closed, the set $\mathbb{R}^N \setminus p^{-1}(0)$ is generic.*

Proof. We refer the reader to a short, elegant proof by Richard Caron and Tim Traynor in an unpublished note [9].

□

Let us define the space M^n of real $(n \times n)$-matrices,

$$M^n = \{A = (a_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,n}}\} \equiv \mathbb{R}^{n \times n} \ .$$

A "nice" property for a matrix A in M^n is, *e.g.*, to be nonsingular or equivalently to satisfy $\det(A) \neq 0$. In Theorem 1.1 we have already shown that

$$M^n_r = \{A \in M^n \mid \det(A) \neq 0\} \quad \text{is a generic subset of } M^n \ . \tag{3.1}$$

This statement is generalized in

EX. 3.1. *Let $M^{n,m}$ denote the space of real $(n \times m)$-matrices, $M^{n,m} \equiv \mathbb{R}^{n \times m}$. Assume $m \geq n$. Show that the set $M^{n,m}_r := \{A \in \mathbb{R}^{n,m} \mid \text{rank } (A) = n\}$ is a generic set in $\mathbb{R}^{n \times m}$.*

Proof. Decompose the matrix $A \in M^{n,m}$, $m \geq n$, as

$$A = (A^1, A^2) \ , \quad A^1 \in M^{n,n} \ , \quad A^2 \in M^{n,m-n} \ .$$

Then by Theorem 1.1 the set $S_1 = \{A^1 \in \mathbb{R}^{n \times n} \mid \det(A^1) \neq 0\}$ is generic in $\mathbb{R}^{n \times n}$. In view of Lemma 2.3 also the set $S = \{(A^1, A^2) \in \mathbb{R}^{n \times m} \mid \det(A^1) \neq 0\}$ is generic in $\mathbb{R}^{n \times m}$. But clearly $S \subset M^{n,m}_r$.

□

The next corollary shows that generically a system $Ax = b$ with $A \in M^n$, $b \in \mathbb{R}^n$, has a unique solution x satisfying $x_i \neq 0$ for all $i = 1, \dots, n$.

COROLLARY 3.1. *In the problem set $\mathcal{P} := \{(A, b) \mid A \in M^n, b \in \mathbb{R}^n\} \equiv \mathbb{R}^{n \times n + n}$ let us consider the set*

$$\mathcal{P}_r := \{(A, b) \in \mathcal{P} \mid Ax = b \textit{ has a unique solution satisfying } x_i \neq 0, \forall i = 1, \dots, n\} \ .$$

Then the set $\mathcal{P}_r$ is a generic subset of $\mathcal{P}$.

Proof. Recall Cramer's rule for a solution x of $Ax = b$ (valid if $\det(A) \neq 0$, see (7.3)):

$$x = A^{-1}b \quad \text{with components} \quad x_i = \frac{\det(A_i(b))}{\det(A)}$$

where $A_i(b)$ is the matrix obtained from A by replacing the i-th column $A_{\cdot i}$ of A by b, i.e., $A_i(b)$ is a (general) matrix in $\mathbb{R}^{n\times n}$. We define the sets

$$S_0 := \{(A,b) \in \mathcal{P} \mid \det(A) \neq 0\}, \quad S_i := \{(A,b) \in \mathcal{P} \mid \det(A_i(b)) \neq 0\},\ i = 1,\dots,n\ .$$

By Theorem 1.1 each of these sets is a generic subset of $\mathcal{P}$ and obviously

$$\mathcal{P}_r = \cap_{i=0}^{n} S_i$$

holds, so that in view of Corollary 2.1 also $\mathcal{P}_r$ is generic in $\mathcal{P}$.

□

3.2. Genericity results in linear programming

In this section we apply the genericity results of Section 3.1 to *Linear Programming Problems* (LP). In particular we show that in the generic situation linear programs are "nondegenerate" with respect to the Simplex method.

Let in the sequel $m, n \in \mathbb{N}$ be given, $m, n \geq 1$. Consider the following pair of *primal/dual linear programs*,

$$\begin{array}{lll} (P): & \max c^T x \quad s.t. & x \in \mathcal{F}_P := \{x \in \mathbb{R}^n \mid Ax \leq b\} \\ (D): & \min\ b^T y \quad s.t. & y \in \mathcal{F}_D := \{y \in \mathbb{R}^m \mid A^T y = c,\ \ y \geq 0\}, \end{array}$$

where A is a given real $m \times n$-matrix and $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ are given vectors. The sets $\mathcal{F}_P$, $\mathcal{F}_D$, respectively, are called feasible sets of $(P), (D)$, respectively. (P) is called the primal and (D) the corresponding dual program. Throughout the section the abbreviation $J := \{1, 2, \dots, m\}$ is used. We introduce the optimal values v_P, v_D of $(P), (D)$ by

$$(3.2) \qquad v_P := \sup_{x\in\mathcal{F}_P} c^T x\ , \qquad v_D := \inf_{y\in\mathcal{F}_D} b^T y\ ,$$

and define $v_P := -\infty$ if $\mathcal{F}_P = \emptyset$, $v_D = \infty$ if $\mathcal{F}_D = \emptyset$. In the sequel, $A_{j\cdot},\ j = 1,\dots,m$, denote the rows of A, and $A_{\cdot i},\ i = 1,\dots,n$, the columns.

According to the next remark, wlog. we can assume that A^T has full row rank n. Then we also have $n \leq m$.

REMARK 3.1. In our linear programs $(P), (D)$ we can assume rank $A^T = n$. Indeed suppose that the rows $A^T_{\cdot i}$ of A^T are linearly dependent, e.g., assume that the last row $A^T_{\cdot n}$ is a linear combination of the other rows. Applying a Gauss elimination step, by adding an appropriate combination of the rows $A^T_{\cdot i}$, $i = 1,\dots,n-1$, to $A^T_{\cdot n}$, the last row is transformed to the zero vector. If by the same operation the value c_n is also transformed to zero we can skip the last row of $A^T y = c$. If c_n is transformed to a value $\alpha \neq 0$, then the condition $A^T y = c$ is not feasible and (see Theorem 3.2(b)) (P) is unbounded.

For any pair of $x \in \mathcal{F}_P, y \in \mathcal{F}_D$ we obtain the *weak duality* relation,

$$b^T y - c^T x = y^T(b - Ax) \geq 0 \qquad \text{or} \qquad c^T x \leq b^T y \tag{3.3}$$

implying

$$v_P \leq v_D \ .$$

Note that from (3.3) it follows that if $x \in \mathcal{F}_P, y \in \mathcal{F}_D$ satisfy $c^T x = b^T y$ then both must be optimal solutions of (P), (D), respectively. We summarize the well-known strong duality results in LP.

THEOREM 3.2. *For the programs $(P), (D)$ precisely one of the following three alternatives holds:*

(a) Both programs are infeasible, i.e.*, $v_P = -\infty$ and $v_D = \infty$.*

(b) Precisely one is feasible. In case $\mathcal{F}_P = \emptyset, \mathcal{F}_D \neq \emptyset$ we then have $v_P = v_D = -\infty$, and in case $\mathcal{F}_P \neq \emptyset, \mathcal{F}_D = \emptyset$, $v_P = v_D = \infty$.

(c) Both programs are feasible. Then for both, (P) and (D), optimal solutions $x \in \mathcal{F}_P, y \in \mathcal{F}_D$ exist and $v_P = c^T x = b^T y = v_D$ holds.

Proof. We refer to [13, Section 4.1] for a proof. □

By this duality result, feasible points $x \in \mathcal{F}_P, y \in \mathcal{F}_D$, are optimal solutions if and only if $c^T x = b^T y$ holds, or the equivalent *complementarity condition* is valid:

$$b^T y - c^T x = y^T(b - Ax) = 0 \quad \text{or} \quad y_j(b_j - A_{j\cdot}x) = 0 \ \forall j \in J \ .$$

Such optimal solutions x, y are called *strictly complementary* if,

$$\text{(SC)} \qquad \text{for all } \ j \in J: \qquad y_j = 0 \ \Leftrightarrow \ (b_j - A_{j\cdot}x) > 0$$

Recall the abbreviation $J = \{1, \ldots, m\}$. In the following, for any subset $J_0 \subset J$ we define the subvector y_{J_0} of $y \in \mathbb{R}^m$ and the submatrix A_{J_0} of $A \in \mathbb{R}^{m \times n}$ by

$$y_{J_0} = \begin{pmatrix} y_j \\ \vdots \end{pmatrix}_{j \in J_0} , \qquad A_{J_0} = \begin{pmatrix} A_{j\cdot} \\ \vdots \end{pmatrix}_{j \in J_0} .$$

DEFINITION 3.1. [vertex, nondegenerate vertex]
For $\overline{x} \in \mathcal{F}_P$ we define the active index set $J_{\overline{x}} := \{j \in J \mid A_{j\cdot}\overline{x} = b_j\}$*. A feasible $\overline{x} \in \mathcal{F}_P$ is called primal vertex (or extreme point of $\mathcal{F}_P$) if $A_{J_{\overline{x}}}$ has rank n, implying $|J_{\overline{x}}| \geq n$. The vertex $\overline{x}$ is called* nondegenerate *if*

$$|J_{\overline{x}}| = n, \text{ and thus } A_{J_{\overline{x}}} \text{ is nonsingular with } \overline{x} \text{ as unique solution of } A_{J_{\overline{x}}}x = b_{J_{\overline{x}}} \ . \tag{3.4}$$

In case $|J_{\overline{x}}| > n$, a vertex $\overline{x}$ is said to be degenerate.

A feasible point $\overline{y} \in \mathcal{F}_D$ is called a dual vertex (or basic solution of $\mathcal{F}_D$) if there exists an index set $J_B \subset J$, $|J_B| = n$, such that $A^T_{J_B}$ is nonsingular (such a submatrix exists since A^T has rank n) and the vector $\overline{y}_{J_B} \geq 0$ solves

$$A^T_{J_B}\overline{y}_{J_B} = c \ . \tag{3.5}$$

So, the dual vertex $\overline{y}$ is given componentwise by $\overline{y}_j,\ j \in J_B$, and $\overline{y}_j = 0,\ j \in J \setminus J_B$. The dual vertex $\overline{y}$ is said to be nondegenerate *if $\overline{y}_{J_B}$ satisfies*

$$\overline{y}_j > 0\ , \quad \textit{for all}\ \ j \in J_B\ . \tag{3.6}$$

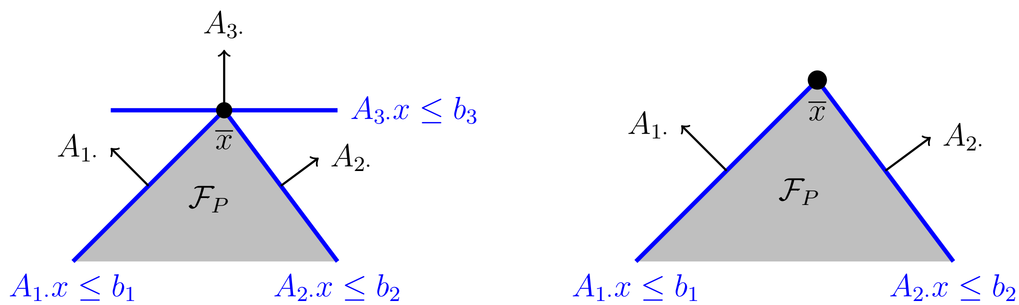

FIGURE 3.1. Primal vertex $\overline{x}$ (left) and nondegenerate vertex $\overline{x}$ (right).

We summarise some facts on vertex solutions.

LEMMA 3.1.
(a) Let $v_P = v_D$ be finite, i.e., optimal solutions $\overline{x}, \overline{y}$ of $(P), (D)$ exist.
Then if F_P has a vertex, the max value of (P) is (also) attained at some vertex.
In this case, $F_D \neq \emptyset$, a dual vertex always exists, and the min value is (also) attained at some dual vertex.

(b) If there exists a nondegenerate optimal primal vertex $\overline{x} \in \mathcal{F}_P$ then (D) has a unique optimal solution $\overline{y}$ (and $\overline{y}$ is a vertex).

(c) If there is a nondegenerate optimal dual vertex $\overline{y} \in \mathcal{F}_D$ then (P) has a unique optimal solution $\overline{x}$ (and $\overline{x}$ is a vertex).

(d) If $\overline{x}, \overline{y}$ is a pair of nondegenerate optimal vertex solutions of $(P), (D)$, then strict complementarity SC holds.

Proof. (a) We refer to [4, Theorem 2.7].
(b) Recall that for a pair $\overline{x}, \overline{y}$ of optimal solutions the complementarity conditions

$$y_j^T (b - Ax)_j = 0\ , \quad j = 1, \ldots, m\ , \tag{3.7}$$

hold. Let now $\overline{x}$ be an optimal nondegenerate vertex solution, *i.e.*, we have $|J_{\overline{x}}| = n,\ A_{J_{\overline{x}}}$ is nonsingular, and

$$(b - Ax)_{J_{\overline{x}}} = 0, \quad (b - Ax)_j > 0\ \forall j \in J_{\overline{x}}^c := J \setminus J_{\overline{x}}\ .$$

According to (3.7) for any optimal solution $\overline{y}$ it follows

$$\overline{y}_j = 0,\ j \in J_{\overline{x}}^c \qquad \text{and} \qquad [A_{J_{\overline{x}}}]^T \overline{y}_{J_{\overline{x}}} = c\ ,$$

which uniquely determines $\overline{y}$, and $\overline{y}$ is a dual vertex with $J_B = J_{\overline{x}}$.

(c) Let $\overline{y}$ be an optimal nondegenerate vertex solution of (D). Then with the corresponding index set J_B we have

$$|J_B| = n, \;\; \overline{y}_j > 0, \; j \in J_B \;, \quad \text{and} \quad [A_{J_B}]^T \quad \text{is nonsingular} \;.$$

In view of (3.7) any primal optimal solution $\overline{x}$ must satisfy

$$A_{J_B}\overline{x} = b_{J_B}$$

which uniquely determines $\overline{x}$. Since $J_B \subset J_{\overline{x}}$ and A_{J_B} is nonsingular, $\overline{x}$ is a vertex.

(d) If both optimal solutions $\overline{x}, \overline{y}$ are nondegenerate (thus unique) by the arguments in (b),(c) we must have $|J_B| = n = |J_{\overline{x}}|$, and $J_B \subset J_{\overline{x}}$ implies $J_B = J_{\overline{x}}$. Together with the nondegeneracy condition this is equivalent to strict complementarity.

□

By definition, each primal vertex $\overline{x}$ is determined by a subset $J_0 \subset J_{\overline{x}}$ with $|J_0| = n$ (such that A_{J_0} is nonsingular). A dual vertex $\overline{y}$ is determined by the index set $J_B, \; |J_B| = n$ (such that A_{J_B} is nonsingular). By noticing that there are $\binom{m}{n}$ different possibilities to select n elements out of $J = \{1, \ldots, m\}$ the following is immediate.

LEMMA 3.2. *The number of vertices of (P) and of (D) is less than or equal to $\binom{m}{n}$.*

A common method to solve linear programs is the

Simplex method. The primal version of this method proceeds as follows:

> Starting from a vertex x_0 of $\mathcal{F}_P$ the algorithm generates a sequence of primal vertices $x_k, k = 1, 2, \ldots$ (or stops with an unbounded direction). If the vertex x_k is nondegenerate, then this algorithm guaranties [31, 13].
>
> $$c^T x_k < c^T x_{k+1} \;.$$
>
> If this holds at all generated vertices x_k, then all vertices x_k must be distinct, and since there are finitely many vertices, the algorithm must end up after at most $\binom{m}{n}$ steps with an optimal vertex (or with an unbounded direction implying that the problem is unbounded). The same holds for the dual version.
>
> Note that at a degenerate vertex a so-called cycling can occur in the Simplex method and the method can generate infinitely many times the same non-optimal vertex. In this case, the Simplex method does not find the solution of (P). For more details on Linear Programming the reader is referred to, e.g., [31, 13].

We are now going to analyse whether generically we can exclude the appearance of degenerate vertices in linear programming and thus generically exclude a cycling during the simplex algorithm.

We first have to define the problem set properly. For given n and m, $m \geq n$, a pair of problems $(P), (D)$ can be seen as an element of the space

$$\mathcal{P}^{m,n} = \{(A, b, c) \mid A \in M^{m,n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n\} \equiv \mathbb{R}^N, \; N := n \cdot m + m + n \;.$$

Recall that a subset $S \subset \mathcal{P}^{m,n}$ is called generic (see Definition 1.2) if S contains a set S_r such that $\mathcal{P}^{m,n} \setminus S_r$ has measure zero, and S_r is open. We will first consider problem (P).

LEMMA 3.3. *There is a generic subset S_P of $\mathcal{P}^{m,n}$ such that for all instances $(A, b, c) \in S_P$ all vertices of the corresponding primal program (P) are nondegenerate.*

Proof. By definition, a vertex $\overline{x}$ of $\mathcal{F}_P$ is degenerate if $|J_{\overline{x}}| \geq n+1$. In view of $A_{J_{\overline{x}}}\overline{x} = b_{J_{\overline{x}}}$, this implies that the

$$|J_{\overline{x}}| \times (n+1)\text{-matrix} \quad (A_{J_{\overline{x}}}, b_{J_{\overline{x}}}) \quad \text{has rank} < (n+1)\ .$$

But in view of Ex. 3.1, for a generic subset $S(J_{\overline{x}})$ of $\mathcal{P}^{m,n}$ this is excluded. There are only finitely many such subsets $J_{\overline{x}} \subset J$. So in view of Corollary 2.1 the set

$$S_P := \cap_{J_{\overline{x}} \subset J} S(J_{\overline{x}})$$

is a generic set in $\mathcal{P}^{m,n}$. By definition, for all instances $(A, b, c) \in S_P$ all vertices of the corresponding primal program (P) are nondegenerate.

□

We obtain the same result for the dual program.

LEMMA 3.4. *There is a generic subset S_D of $\mathcal{P}^{m,n}$ such that for all instances $(A, b, c) \in S_D$ all vertices of the corresponding dual program (D) are nondegenerate.*

Proof. A vertex $\overline{y}$ of $\mathcal{F}_D$ is degenerate if with the corresponding (basic) index set J_B $|J_B| = n$, the solution $\overline{y}_{J_B} \in \mathbb{R}^n$ of

$$[A_{J_B}]^T \overline{y}_{J_B} = c \quad \text{satisfies } \overline{y}_{j_0} = 0 \quad \text{for some } j_0 \in J_B\ .$$

But by Corollary 3.1, for a generic subset $S(J_B)$ of $\mathcal{P}^{m,n}$ this is excluded. Since there are only finitely many such subsets $J_B \subset J$, according to Corollary 2.1, the set

$$S_D := \cap_{J_B \subset J} S(J_B)$$

is a generic set in $\mathcal{P}^{m,n}$. By definition, for all instances $(A, b, c) \in S_D$ all vertices of the corresponding dual program (D) are nondegenerate.

□

By defining

$$\mathcal{P}_r^{m,n} = S_P \cap S_D\ ,$$

from Lemma 3.3 and Lemma 3.4 we conclude that the set $\mathcal{P}_r^{m,n}$ is generic in $\mathcal{P}^{m,n}$.

COROLLARY 3.2. *There is a generic subset $\mathcal{P}_r^{m,n} \subset \mathcal{P}^{m,n}$ such that for all instances (A, b, c) in $\mathcal{P}_r^{m,n}$ all primal and dual vertices of the corresponding programs $(P), (D)$ are nondegenerate.*

Corollary 3.2 implies that, given a "degenerate problem" $(\overline{A}, \overline{b}, \overline{c})$, by (almost all) arbitrarily small perturbations we obtain a "nondegenerate problem" (A, b, c), i.e., a problem $(A, b, c) \in \mathcal{P}_r^{m,n}$. This explains the fact, that in practice the Simplex method works well even without any anticycling procedure.

We add some remarks on feasibility. We emphasize, that the generic set $\mathcal{P}_r^{n,m}$ in Corollary 3.2 does contain problem pairs such that (P) or (D) are infeasible. We give a simple example in $\mathcal{P}_r^{2,1}$:

$$\begin{aligned} (P): \quad & \max_{x\in\mathbb{R}} \; x & \text{s.t.} \quad & x \le -1 \\ & & & -x \le -1 \\ (D): \quad & \min_{y\in\mathbb{R}^2} -y_1 - y_2 & \text{s.t.} \quad & y_1 - y_2 = 1 \\ & & & y_1, y_2 \ge 0 \end{aligned}$$

Here, $\mathcal{F}_P = \emptyset$, i.e., (P) is infeasible and (D) possesses one vertex $\overline{y} = (1,0)$, which is nondegenerate. Note, that beginning with $\overline{y} = (1,0)$ the (dual) Simplex-method immediately finds the unbounded edge $y(t) = \binom{1}{0} + t\binom{1}{1}$, $t \ge 0$, showing that (D) is unbounded. These properties are stable wrt. small perturbations in (A,b,c).

We also give an example in $\mathcal{P}_r^{2,2}$ with $\mathcal{F}_P = \emptyset$, and $\mathcal{F}_D = \emptyset$:

$$\begin{aligned} (P): \quad & \max_{x\in\mathbb{R}^2} \; x_1 + 2x_2 & \text{s.t.} \quad & x_1 + x_2 \le \;\; 1 \\ & & & -x_1 - x_2 \le -2 \\ (D): \quad & \min_{y\in\mathbb{R}^2} y_1 - 2y_2 & \text{s.t.} \quad & \binom{1}{1} y_1 + \binom{-1}{-1} y_2 = \binom{1}{2} \\ & & & y_1, y_2 \ge 0 \end{aligned} \tag{3.8}$$

We emphasize that the condition $\mathcal{F}_P = \mathcal{F}_D = \emptyset$ is not stable wrt. to small perturbations in the whole parameter set (A,b,c) (see [12]). Also in our example, almost all small perturbations of A make the program (P) in (3.8) feasible (see, Figure 3.2).

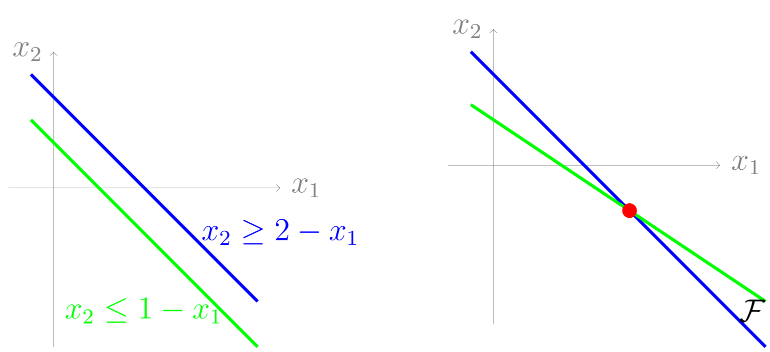

FIGURE 3.2. Empty feasible set (left) and perturbed nonempty feasible set (right).

We add some comments on topological properties of, *e.g.*, the set of instances $(A,b,c) \in \mathcal{P}^{m,n}$ such that (P) is feasible.

LEMMA 3.5.
(a) The set $S_1 := \{(A,b) \in \mathbb{R}^{m\times(n+1)} \mid Ax \le b$ *is feasible*$\}$ *is neither closed nor open. So also the complement set of instances which are infeasible are neither open nor closed.*
(b) The set $S_0 := \{(A,b) \in \mathbb{R}^{m\times(n+1)} \mid Ax < b$ *has a solution*$\}$ *is open and* $S_0 = \operatorname{int}(S_1)$.

Proof. (a) We give the proof by counterexamples. To show non-closedness we consider the sequence of feasible systems in $\mathbb{R}^2$:

$$x_1 \geq 1 \ , \quad x_1 - \frac{1}{k} x_2 \leq 0 \text{ for } \ k \in \mathbb{N} \ .$$

Obviously, the points $(1, k), k \in \mathbb{N}$ are feasible. For $k \to \infty$ the limiting system $x_1 \geq 1, \ x_1 \leq 0$, is however infeasible.

An example that shows non-openness is given by the following inequalities in $\mathbb{R}$: The system $x_1 \leq 0, x_1 \geq 0$ is feasible but for any $\varepsilon > 0$ the system $x_1 \leq 0, x_1 \geq \varepsilon$ is not.

(b) The set S_0 is obviously open. Indeed, if $\overline{A}\overline{x} < \overline{b}$ holds then $A\overline{x} < b$ for all (A, b) near $(\overline{A}, \overline{b})$. To prove $S_0 = \text{int}\,(S_1)$ it is sufficient to show that an instance $(\overline{A}, \overline{b}) \in S_1 \setminus S_0$ is a boundary point of S_1. To do so, we show that given an instance $(\overline{A}, \overline{b}) \in S_1 \setminus S_0$, by an appropriate small perturbation the instance becomes infeasible. Such a pertubation is given, *e.g.*, by $(A, b) := (\overline{A}, \overline{b} - \varepsilon e), \varepsilon > 0$ ($e = (1, \ldots, 1)$). Indeed, if $\overline{A}x \leq \overline{b} - \varepsilon e$ has a feasible point $\overline{x}$ then $\overline{A}\overline{x} < \overline{b}$ holds, a contradiction.

□

CHAPTER 4

Genericity results in analysis

We are interested in generic properties of (non-parametric) problems in analysis, such as nonlinear equations, or unconstrained and constrained programs. We have to mention that much of the material in Section 4.3 and Section 4.5 is essentially taken from the book [21].

4.1. Basic results

In this section we sketch basic results from analysis needed to prove the genericity results lateron.

Partition of unity. Recall that the support supp f of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined as (the closed set)

$$\text{supp}\, f = \text{cl}\, \{x \in \mathbb{R}^n \mid f(x) \neq 0\}\ .$$

DEFINITION 4.1. *Let $\{U_i\}_{i \in I}$ be an open cover of $\mathbb{R}^n$, with I an arbitrary (countable) index set. A collection $\{\phi_i\}_{i \in I}$ of functions $\phi_i \in C^\infty(\mathbb{R}^n, \mathbb{R})$, $\phi_i(x) \geq 0\ \forall x \in \mathbb{R}^n,\ i \in I$, is called a C^∞-partition of unity subordinate to $\{U_i\}_{i \in I}$ if*

(1) $\text{supp}\, \phi_i \subset U_i$
(2) $\{\text{supp}\, \phi_i\}_{i \in I}$ *is locally finite, i.e., for each $x \in \mathbb{R}^n$ there exists an open neighborhood U_x such that $\{i \in I \mid \text{supp}\, \phi_i \cap U_x \neq \emptyset\}$ is finite.*
(3) $\sum_{i \in I} \phi_i(x) = 1$ *for all* $x \in \mathbb{R}^n$.

EX. 4.1. *Show that condition (3) of this definition implies that also the set*

$$\{\text{int supp}\, \phi_i\}_{i \in I} \text{ is an open cover of } \mathbb{R}^n\ .$$

The following theorem assures the existence of a partition of unity and will be used regularly to globalize a local behavior.

THEOREM 4.1. *Every open cover $\{U_i\}_{i \in I}$ of $\mathbb{R}^n$ has a C^∞-partition of unity subordinate to $\{U_i\}_{i \in I}$.*

Proof. See, e.g., [18, 2.1 Theorem].

□

Concering Definition 4.1 and Theorem 4.1 we have to add an important

REMARK 4.1. We recall some notions from differential geometry. A topological vector space V is called a Hausdorff space if for any pair $x, y \in V$, $x \neq y$, there exist neighborhoods U_x of x and U_y of y satisfying $U_x \cap U_y = \emptyset$. V is said to be paracompact if every open cover of V has an open refinement that is locally finite. The topological vector space

V is called second countable if it has a countable base, i.e., if there is a countable family $\{U_i\}_{i\in\mathbb{N}}$ of open subsets U_i of V such that any open subset U of V can be written as union of a subset of the family $\{U_i\}_{i\in\mathbb{N}}$.
It is well-known that the space $\mathbb{R}^n$ endowed with the standard topology (open Euclidean ε-balls around points) is a paracompact, second countable Hausdorff space. Further, for a locally Euclidean Hausdorff space V the following equivalence holds (see, e.g., [30, Exercise 2.15]):

$$V \text{ is second countable} \quad \Leftrightarrow \quad V \text{ is paracompact with countably many components}$$

We emphasize that in the proof of a C^∞-partition of unity in Theorem 4.1 the paracompactness of the space $\mathbb{R}^n$ (i.e., the existence of a locally finite refinement of $\{U_i\}_{i\in I}$) plays a crucial role.

We regularly will make use of the following result.

LEMMA 4.1. *Let $W \subset \mathbb{R}^n$ be a closed set and let $U,\ W \subset U$, be an open neighborhood of W. Let further $f \in C^k(U, \mathbb{R}^m)$, $k \geq 1$, be given.*
(a) Then there exists a function $\xi \in C^\infty(\mathbb{R}^n, \mathbb{R})$, such that

$$0 \leq \xi(x) \leq 1\ \forall x \in \mathbb{R}^n, \quad \xi(x) = 1 \text{ on an open neighborhood } U_1 \text{ of } W, \quad \operatorname{supp} \xi \subset U\ .$$

(b) There exists $\tilde{f} \in C^k(\mathbb{R}^n, \mathbb{R}^m)$ such that f and $\tilde{f}$ coincide on a neighborhood U_1 of W. Moreover, if W is compact, then $\tilde{f}$ can be chosen such that supp $\tilde{f}$ *is also compact and* supp $\tilde{f} \subset U$.

Proof. See [21, Lemma 2.2.2]. □

Parametric Sard Theorem. This theorem is the basis for the density part of the genericity results in this chapter.

Let be given a function $f \in C^1(\mathbb{R}^n, \mathbb{R}^s)$. Then $0 \in \mathbb{R}^s$ is called a regular value of f if

$$\nabla f(x) \quad \text{has (full) rank } s \text{ for all } x \text{ such that } f(x) = 0\ . \tag{4.1}$$

The following is a useful generalisation of the Sard Theorem (see *e.g.*, [40, Prop. 78.10] for a proof).

THEOREM 4.2. [Parametric Sard Theorem]
Let $h : Q \times P \subset \mathbb{R}^q \times \mathbb{R}^p \to \mathbb{R}^s$, $(x, y) \mapsto h(x, y)$, be a C^k-mapping with $k > \max\{0, q - s\}$ and with open sets $Q \subset \mathbb{R}^q, P \subset \mathbb{R}^p$. If $0 \in \mathbb{R}^s$ is a regular value of h, then for almost all $y \in P$ the value 0 is a regular value of the function $h_1(x) := h(x, y)$.

Note for the case $s > q$, in Theorem 4.2, that then the fact that 0 is a regular value of $h_1(x) = h(x, y)$ (for fixed y) implies that the equation $h_1(x) = 0$ cannot possess any solution.

4.2. Genericity results for eigenvalues of real symmetric matrices

In this section we present some genericity results concerning eigenvalues of real valued matrices. Clearly, real matrices may have complex eigenvalues. However in this report we only are interested in real eigenvalues and corresponding real eigenvectors x. So here we restrict ourselves to symmetric matrices, i.e., matrices from $S^n = \{A \in M^n \mid A = A^T\}$ (see Chapter 7). To find eigenpairs (x, λ), of $A \in S^n$ we can consider solutions $x \in \mathbb{R}^n,\ \lambda \in \mathbb{R}$, of the system:

$$H(x, \lambda; A) := \begin{array}{ll} Ax - \lambda x & = 0 \\ x^T x - 1 & = 0 \end{array} . \tag{4.2}$$

We say that $\lambda \in \mathbb{R}$ is an s-fold eigenvalue of A, if (4.2) allows s linearly independent solutions (eigenvectors) $x_1, \ldots, x_s \in \mathbb{R}^n$ wrt. the same value λ.
The eigenvalue λ is called simple if λ is not a s-fold eigenvalue with $s \geq 2$. Note that if λ is a double eigenvalue of A then with corresponding linearly independent eigenvectors x_1, x_2, for any vector

$$x(\rho) := \frac{x_1 + \rho(x_2 - x_1)}{\|x_1 + \rho(x_2 - x_1)\|} , \quad \rho \in \mathbb{R} ,$$

the pair $(x(\rho), \lambda)$ is a solution of (4.2) (an eigenpair of A). Consequently, an eigenvalue $\overline{\lambda}$ with corresponding (normalized) eigenvector $\overline{x}$ is a simple eigenvalue if

$$\text{the solution} \quad (\overline{x}, \overline{\lambda}) \text{ of (4.2) is locally unique} . \tag{4.3}$$

By the Inverse Function Theorem 7.1 this condition holds if the Jacobian of (4.2) at $(\overline{x}, \overline{\lambda})$ (A fixed),

$$\nabla_{(x,\lambda)} H(\overline{x}, \overline{\lambda}; A) = \begin{pmatrix} A - \overline{\lambda} I & -\overline{x} \\ 2\overline{x}^T & 0 \end{pmatrix} \text{ is nonsingular} . \tag{4.4}$$

By applying the Parametric Sard Theorem 4.2 we now prove that the following (nice) property is generic for $A \in S^n$:

Property. All eigenvalues of $A \in S^n$ are simple.

THEOREM 4.3. *There is a generic subset $\mathcal{P}_r$ of $\mathcal{P} = S^n$ such that for any $A \in \mathcal{P}_r$ all eigenvalues λ of A are simple.*

Proof. weakly generic: To apply Theorem 4.2, we have to show that at all solutions $(x, \lambda; A)$ of $H(x, \lambda; A) = 0$ in (4.2) the Jacobian

$$\nabla_{(x,\lambda;A)} H(x, \lambda; A) = \left(\begin{array}{cc|c} A - \lambda I & -x & \nabla_A (Ax - \lambda x) \\ 2x^T & 0 & 0 \end{array} \right)$$

has full row rank $n+1$. Due to symmetry, the matrix A has the form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{12} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{13} & a_{23} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \dots & a_{nn} \end{pmatrix} ,$$

and the Jacobian $\nabla_A(Ax - \lambda x) = \nabla_A Ax$ reads,

$$\begin{pmatrix} \overbrace{x_1}^{a_{11}} & \overbrace{x_2}^{a_{12}} & \overbrace{x_3}^{a_{13}} & \dots & \overbrace{x_n}^{a_{1n}} & \overbrace{0}^{a_{22}} & \overbrace{0}^{a_{23}} & \dots & \dots & \overbrace{0}^{a_{33}} & \dots & \dots & \overbrace{0}^{a_{nn}} \\ 0 & x_1 & 0 & \dots & 0 & x_2 & x_3 & \dots & \dots & 0 & \dots & \dots & 0 \\ 0 & 0 & x_1 & \dots & 0 & 0 & x_2 & \dots & \dots & x_3 & \dots & \dots & 0 \\ \vdots & \dots & \dots & \ddots & \dots & & & & & & & \dots & \vdots \\ 0 & 0 & 0 & \dots & x_1 & \dots & & & & & & \dots & x_n \end{pmatrix} .$$

A moment of reflection shows that in view of $x \neq 0$ this matrix has full row rank n. Indeed, if $x_1 \neq 0$ this is directly clear. The same holds for $x_2 \neq 0$, or $x_3 \neq 0$, etc. So by Theorem 4.2, for almost all $A \in S^n$, i.e., for a weakly generic subset $\mathcal{P}_r \subset S^n$ the following is true: for all $A \in \mathcal{P}_r$ at any solution (x, λ) of (4.2) the Jacobian wrt. (x, λ),

$$\nabla_{(x,\lambda)} H(x, \lambda; A) = \begin{pmatrix} A - \lambda I & -x \\ 2x^T & 0 \end{pmatrix} \quad \text{has full row rank } n+1 .$$

By the arguments above (see (4.4)), for all $A \in \mathcal{P}_r$ all eigenvalues of A are simple.

<u>*openness*</u>: Take $\overline{A} \in \mathcal{P}_r$ arbitrarily and consider all its (real) simple eigenvalues $\overline{\lambda}_1 < \ldots < \overline{\lambda}_n$, with corresponding (orthonormal) eigenvectors $\overline{x}_i$, $i = 1, \ldots, n$, such that the Jacobians

$$\begin{pmatrix} \overline{A} - \overline{\lambda}_i I & -\overline{x}_i \\ 2\overline{x}_i^T & 0 \end{pmatrix} \quad \text{are nonsingular} \;, \; i = 1, \ldots, n .$$

According to the Implicit Function Theorem 7.2, to any eigenpair $(\overline{x}_i, \overline{\lambda}_i)$ there exists $\varepsilon_i > 0$, neigborhoods $B(\overline{A}, \varepsilon_i)$, and C^1-functions $(x_i, \lambda_i) : B(\overline{A}, \varepsilon_i) \to \mathbb{R}^{n+1}$ with $x_i(\overline{A}) = \overline{x}_i, \lambda_i(\overline{A}) = \overline{\lambda}_i$, such that for $A \in B(\overline{A}, \varepsilon_i)$:

$$\big(x_i(A), \lambda_i(A)\big) \quad \text{are locally unique (isolated) eigenpairs of } A .$$

Then for $0 < \varepsilon \leq \min\{\varepsilon_i \mid i = 1, \ldots, n\}$, ε small, the n eigenvalues $\lambda_i(A)$, $A \in B(\overline{A}, \varepsilon)$, still satisfy

$$\lambda_1(A) < \ldots < \lambda_n(A) ,$$

with corresponding (orthonormal) eigenvectors $x_i(A)$, i.e., the eigenvalues $\lambda_i(A)$ are simple.

□

4.3. Genericity results for unconstrained programs

For a given function $f \in C^2(\mathbb{R}^n, \mathbb{R})$ we consider the unconstrained optimization problem to find local minimizers of f,

$$(P_0): \qquad \min_{x \in \mathbb{R}^n} \; f(x)$$

It is well-known that at each local minimizer $\overline{x}$ the first and second order optimality conditions

$$\nabla f(\overline{x}) = 0 \quad \text{and} \quad \nabla^2 f(\overline{x}) \; \text{ is positive semidefinite}$$

must hold. These conditions are necessary but not sufficient for $\overline{x}$ to be a local minimizer. We however have

LEMMA 4.2. *Let $f \in C^2(\mathbb{R}^n, \mathbb{R})$ and $\overline{x} \in \mathbb{R}^n$ such that*

$$\nabla f(\overline{x}) = 0 \quad \textit{and} \quad \nabla^2 f(\overline{x}) \; \textit{ is positive definite} \; .$$

Then $\overline{x}$ is an isolated, strict local minimizer of f.

Proof. See, e.g., [13, Lemma 11.2, Ex. 11.4].

□

To solve (P_0) so-called Newton type methods can be used. The classical version of such a method computes critical points , i.e., solutions $\overline{x}$ of

$$\nabla f(\overline{x}) = 0 \; ,$$

via the following Newton procedure: Chose a starting point x_0 and iterate

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla^T f(x_k) \; , \qquad k = 0, 1, \ldots \; . \tag{4.5}$$

Under the condition that the Hessian $\nabla^2 f(\overline{x})$ is nonsingular at a critical point $\overline{x}$, this iteration is locally quadratically convergent.

THEOREM 4.4. [Convergence result for the Newton iteration]
Let $f \in C^3(\mathbb{R}^n, \mathbb{R})$ and $\overline{x} \in \mathbb{R}^n$ such that

$$\nabla f(\overline{x}) = 0 \qquad \textit{and} \quad \nabla^2 f(\overline{x}) \quad \textit{is nonsingular} \; .$$

Then there exists a neighborhood $U_{\overline{x}}$ of $\overline{x}$ such that starting with any $x_0 \in U_{\overline{x}}$ the proceedure (4.5) *converges quadratically to $\overline{x}$, i.e., with some $C > 0$ we have*

$$\|x_{k+1} - \overline{x}\| \le C \; \|x_k - \overline{x}\|^2 \; , \quad k = 0, 1, \ldots \; .$$

Proof. We refer to, e.g., [13, Theorem 11.4] for a proof.

□

We will call a point $\overline{x}$ satisfying $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x})$ nonsingular, a nondegenerate critical point.
For the genericity results, now, we restrict ourselves to function f in $C^\infty(\mathbb{R}^n, \mathbb{R})$. We however mention that all results remain true for functions f in $C^k(\mathbb{R}^n, \mathbb{R})$, $k \ge 2$.
According to the above facts, with regard to the Newton method for solving the minimation problem (P_0), the following property would be desirable for $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$.

Property. The function f belongs to the set

$$\mathcal{P}_r := \big\{ f \in C^\infty(\mathbb{R}^n, \mathbb{R}) \mid \quad \text{at all solutions } \overline{x} \in \mathbb{R}^n \text{ of } \nabla f(\overline{x}) = 0 \text{ the Hessian } \nabla^2 f(\overline{x}) \text{ is nonsingular} \big\} \ . \tag{4.6}$$

In the sequel we will show that this set $\mathcal{P}_r$ is open and dense and thus generic wrt. the C_s^k-topology (cf., Definition 2.2), $k \geq 2$.

The proof of the density part will be based on the following lemma.

LEMMA 4.3. *Let $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$ be given arbitrarily. Then for almost all $a \in \mathbb{R}^n$ the perturbed function $f_a(x) := f(x) + a^T x$ belongs to $\mathcal{P}_r$.*

Proof. To prove the statement, we will apply the Parametric Sard Theorem 4.2 to the equation

$$H(x; a) := \nabla^T f_a(x) = \nabla^T f(x) + a = 0 \ .$$

The Jacobian of H wrt. to $(x; a)$ has the form (I_n is the unit matrix in M^n)

$$\nabla_{(x;a)} H(x; a) = \big(\ \nabla^2 f(x) \mid I_n \ \big) \ .$$

Obviously, this matrix has full row rank n, and by the Parametric Sard Theorem for almost all $a \in \mathbb{R}^n$ the following is true: at all solutions x of

$$H(x; a) = \nabla^T f_a(x) = 0 \ ,$$

the Jacobian $\nabla_x H(x; a) = \nabla^2 f(x)$ has full row rank n, i.e., $\nabla^2 f(x)$ is nonsingular.

□

We are now ready to prove the main genericity statement in unconstrained programming.

THEOREM 4.5. [Genericity result for unconstrained programs]
The set $\mathcal{P}_r$ in (4.6) *is open and dense in $C^\infty(\mathbb{R}^n, \mathbb{R})$ wrt. the C_s^k-toloplogy for any $k \geq 2$. This implies that $\mathcal{P}_r$ is also C_s^k-dense for $k = 0, 1$ (see Remark 2.1).*

Proof. The idea of the proof is taken from the proofs of Theorem 7.1.13 and Theorem 7.3.10 in [21]. We start by defining the function

$$\sigma_f(x) := \|\nabla f(x)\| + |\det \nabla^2 f(x)| \ , \quad x \in \mathbb{R}^n \ .$$

This function obviously has the following property for $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$:

$$f \in \mathcal{P}_r \quad \Leftrightarrow \quad \sigma_f(x) > 0 \quad \forall x \in \mathbb{R}^n \ .$$

openness part. Let $f \in \mathcal{P}_r$, i.e.,

$$\sigma_f(x) > 0 \quad \text{ for all } x \in \mathbb{R}^n \ .$$

Take any $\overline{x} \in \mathbb{R}^n$ (arbitrarily). Then by defining $B(\overline{x}, \delta) := \{x \in \mathbb{R}^n \mid \|x - \overline{x}\| < \delta\}$ we have,

$$\eta_{\overline{x}} := \min_{x \in \text{cl } B(\overline{x},1)} \sigma_f(x) > 0 \ ,$$

and there exists $\varepsilon_{\overline{x}} > 0$, such that for any $g \in C^\infty(\mathbb{R}^n, \mathbb{R})$ satisfying

$$\|f - g\|_{k, \text{cl } B(\overline{x},1)} < \varepsilon_{\overline{x}}$$

(cf., (2.6)) we have

$$\min_{x \in \text{cl}\, B(\overline{x},1)} \sigma_g(x) > \frac{1}{2}\eta_{\overline{x}} \; . \tag{4.7}$$

We now can chose a sequence of points $x_i \in \mathbb{R}^n$, $i \in \mathbb{N}$, such that the open balls $\{B(x_i,1)\}_{i\in\mathbb{N}}$ cover $\mathbb{R}^n$ and that this cover is locally finite. Let further $\{\theta_i\}_{i\in\mathbb{N}}$ be a partition of unity subordinate to $\{B(x_i,1)\}_{i\in\mathbb{N}}$ (cf., Definition 4.1) and let $\varepsilon_{x_i}, \eta_{x_i}$ be defined as above: for all $i \in \mathbb{N}$ we have

$$\min_{x \in \text{cl}\, B(x_i,1)} \sigma_g(x) > \frac{1}{2}\eta_{x_i} \quad \forall g \in C^\infty(\mathbb{R}^n,\mathbb{R}) \text{ with } \|f-g\|_{k,\text{cl}\, B(x_i,1)} < \varepsilon_{x_i} \; .$$

Since $\{B(x_i,1)\}_{i\in\mathbb{N}}$ is a locally finite cover (cf., Definition 4.1), for any $i_0 \in \mathbb{N}$ the set $I(i_0) = \{i \in \mathbb{N} \mid \text{supp}\,\theta_i \cap \text{supp}\,\theta_{i_0} \neq \emptyset\}$ is finite. We define $\overline{\varepsilon}_{i_0} = \min_{i\in I(i_0)} \varepsilon_{x_i}$ and the strictly positive continuous function

$$\rho(x) = \sum_{i\in\mathbb{N}} \overline{\varepsilon}_i \theta_i(x) \; .$$

For an arbitrary point $x_0 \in \text{supp}\,\theta_{i_0}$ we find

$$\rho(x_0) = \sum_{i\in I(i_0)} \overline{\varepsilon}_i \theta_i(x_0) \le \sum_{i\in I(i_0)} \varepsilon_{x_{i_0}} \theta_i(x_0) \le \varepsilon_{x_{i_0}} \; .$$

Consequently, by construction, the function $\rho(x)$ defines a C^k_s-neighborhood $U^k_{\rho,f}$ such that for any $g \in U^k_{\rho,f}$ the function $\sigma_g(x)$ is positive on $\mathbb{R}^n$ and thus g is contained in $\mathcal{P}_r$.

density part. Let $g \in C^\infty(\mathbb{R}^n,\mathbb{R})$ be an arbitrary function and let $U^k_{\phi,g}$ be an arbitrary C^k_s-neigborhood defined by the function $\phi \in C^0_+(\mathbb{R}^n,\mathbb{R})$.
We now construct a sequence of functions $f_i \in C^\infty(\mathbb{R}^n,\mathbb{R}) \cap U^k_{\phi,g}$, $i \in \mathbb{N}$, such that (the pointwise limit)

$$f := \lim_{i\to\infty} f_i \quad \text{ is contained in } \mathcal{P}_r \cap U^k_{\phi,g} \; .$$

Recall the open cover $\{B(x_i,1)\}_{i\in\mathbb{N}}$ of $\mathbb{R}^n$, and the corresponding partition of unity $\{\theta_i\}_{i\in\mathbb{N}}$ in the openness part. By Ex. 4.1 also $\{\text{int}\,\text{supp}\,\theta_i\}_{i\in\mathbb{N}}$ is an open cover of $\mathbb{R}^n$ and we can chose a partition $\{\chi_i\}_{i\in\mathbb{N}}$ of unity subordinate to $\{\text{int}\,\text{supp}\,\theta_i\}_{i\in\mathbb{N}}$. Note that the sets $\text{supp}\,\theta_i$ are closed and satisfy

$$\text{int}\,\text{supp}\,\theta_i \subset \text{supp}\,\theta_i \subset B(x_i,1) \; .$$

Thus we can find functions $\xi_i \in C^\infty(\mathbb{R}^n,\mathbb{R})$ satisfying (see Lemma 4.1(a))

$$0 \le \xi_i(x) \le 1 \;\forall x \in \mathbb{R}^n, \qquad \xi_i(x) = 1 \text{ on an open neighborhood } W_i \text{ of } \text{supp}\,\chi_i,$$
$$\text{supp}\,\xi_i \subset \text{int}\,\text{supp}\,\theta_i \; .$$

We now start our construction with $i = 1$. According to Lemma 4.3 we can chose $u_1 \in \mathbb{R}^n$ (small enough) such that $h_1(x) := g(x) + u_1^T x$ is contained in $\mathcal{P}_r$ and

$$\|\xi_1(x)u_1^T x\|_{k,\text{supp}\,\theta_1} < \varepsilon_1 =: \min_{x\in\text{supp}\,\theta_1} \phi(x) \; . \tag{4.8}$$

By defining

$$f_1(x) := \begin{cases} g(x) + \xi_1(x) u_1^T x & x \in B(x_1, 1) \\ g(x) & x \in \mathbb{R}^n \setminus B(x_1, 1) \end{cases}$$

we have given a function $f_1 \in C^\infty(\mathbb{R}^n, \mathbb{R}) \cap U^k_{\phi,g}$, and f_1 is contained in $\mathcal{P}_r(\text{supp}\, \chi_1)$ where for a given closed $C \subset \mathbb{R}^n$ we put

$$\mathcal{P}_r(C) := \{ f \in C^\infty(\mathbb{R}^n, \mathbb{R}) \mid \det(\nabla^2 f(x)) \neq 0 \ \forall x \in C \ \text{ with } \ \nabla f(x) = 0 \} \ .$$

In the same way we define $f_2 \in C^\infty(\mathbb{R}^n, \mathbb{R}) \cap U^k_{\phi,g} \cap \mathcal{P}_r(\cup_{i=1}^2 \text{supp}\, \chi_i)$ by

$$f_2(x) := \begin{cases} f_1(x) + \xi_2(x) u_2^T x & x \in B(x_2, 1) \\ f_1(x) & x \in \mathbb{R}^n \setminus B(x_2, 1) \end{cases} ,$$

where $u_2 \in \mathbb{R}^n$ is chosen such that $f_2 \in \mathcal{P}_r(\text{supp}\, \chi_2)$ and

$$\|\xi_2(x) u_2^T x\|_{k, \text{supp}\, \theta_2} < \varepsilon_2 =: \min_{x \in \text{supp}\, \theta_2} \phi(x) \ .$$

In case $\text{supp}\, \theta_1 \cap \text{supp}\, \theta_2 \neq \emptyset$, we also can force

$$f_2(x) \in \mathcal{P}_r(C) \qquad \text{for } C = \text{supp}\, \chi_1 \ .$$

With regard to the proof of the openness part of $\mathcal{P}_r$ this is possible by chosing u_2 in the part $\xi_2(x) u_2^T x$ of f_2 small enough.
In this way after s steps we have constructed a function $f_s \in C^\infty(\mathbb{R}^n, \mathbb{R}) \cap U^k_{\phi,g}$,

$$f_s(x) := \begin{cases} f_{s-1}(x) + \xi_s(x) u_s^T x & x \in B(x_s, 1) \\ f_{s-1}(x) & x \in \mathbb{R}^n \setminus B(x_s, 1) \end{cases} ,$$

satisfying $f_s \in \mathcal{P}_r(\text{supp}\, \chi_s)$ and by openness arguments also

$$f_s \in \mathcal{P}_r(\cup_{i=1}^{s-1} \text{supp}\, \chi_i) \ .$$

Since the cover $\{\text{supp}\, \theta_i\}_{i \in \mathbb{N}}$ is locally finite, for any $\overline{x} \in \mathbb{R}^n$ there exists a number $s(\overline{x})$ such that

$$f_{j_1}(\overline{x}) = f_{j_2}(\overline{x}) \qquad \text{for all } j_1, j_2 > s(\overline{x}) \ .$$

Thus, for any $x \in \mathbb{R}^n$ the pointwise limit

$$f(x) := \lim_{i \to \infty} f_i(x)$$

is well defined and by construction $f \in C^\infty(\mathbb{R}^n, \mathbb{R}) \cap U^k_{\phi,g}$. Moreover, since $\{\text{supp}\, \chi_i\}_{i \in \mathbb{N}}$ constitutes a cover of $\mathbb{R}^n$ we also have $f \in \mathcal{P}_r(\mathbb{R}^n) = \mathcal{P}_r$.

□

4.4. Genericity results for systems of nonlinear equations

In Section 3.1 we have shown that matrices $A \in M^{m,n} \equiv \mathbb{R}^{m \times n}$ ($m \leq n$) generically have full rank m. In this section we generalize such a result to functions $F : \mathbb{R}^n \to \mathbb{R}^m$.
For fixed $m, n \in \mathbb{N}$, we consider for vector functions $F(x) = \big(f_1(x), \ldots, f_m(x)\big)^T$ with components $f_j(x) \in C^\infty(\mathbb{R}^n, \mathbb{R})$, $j = 1, \ldots, m$, the system of m equations in n unknowns:

$$(E_0): \qquad F(x) = 0 \qquad \text{or} \qquad f_j(x) = 0 \ , \quad j = 1, \ldots, m \ . \tag{4.9}$$

A desirable property for solving (E_0) is

Property. The function F belongs to

(4.10) $$\mathcal{P}_r := \Big\{ F \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \mid \begin{array}{l} \text{at all solutions } \overline{x} \text{ of } F(x) = 0 \\ \text{the Jacobian } \nabla F(\overline{x}) \text{ has full row rank } m \end{array} \Big\}$$

By using the same technique as in Section 4.3 we will show that the set $\mathcal{P}_r$ is open and dense (i.e., generic) wrt. the C^k_s-topology for any $k \geq 1$. Note that in case $m > n$ the Jacobian ∇F has rank

$$\text{rank } \nabla F(x) \leq \min\{m, n\} = n < m \ .$$

So for $F \in \mathcal{P}_r$ in this case $m > n$, there cannot be any solution of $F(x) = 0$. Again the density proof is based on

LEMMA 4.4. *Let $F \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ be given. Then for almost all $a \in \mathbb{R}^m$ the perturbed function $F(x; a) := F(x) + a$ belongs to $\mathcal{P}_r$.*

Proof. The Jacobian of $F(x; a)$ wrt. to (x, a) has the form

$$\nabla_{(x,a)} F(x; a) = \big(\nabla F(x) \mid I_m \big) \ ,$$

and obviously has full row rank m. So, the Parametric Sard Theorem implies that for almost all $a \in \mathbb{R}^m$ the following is true: at all solutions $\overline{x}$ of

$$F(x; a) = F(x) + a = 0$$

the Jacobian $\nabla_x F(\overline{x}; a) = \nabla F(\overline{x})$ has full row rank m.

□

The genericity statement wrt. problem (E_0) is as follows.

THEOREM 4.6. [Genericity result for systems of equations]
The set $\mathcal{P}_r$ in (4.10) *is open and dense in $C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ wrt. the C^k_s-toloplogy for any $k \geq 1$.*

Proof. As in the proof of Theorem 4.5 we first define a function $\sigma_F(x) \in C^0(\mathbb{R}^n, \mathbb{R})$ such that for $F \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ we have

(4.11) $$F \in \mathcal{P}_r \quad \Leftrightarrow \quad \sigma_F(x) > 0 \ \forall x \in \mathbb{R}^n \ .$$

Obviously such a function is given by

(4.12) $$\sigma_F(x) := \|F(x)\| + \sum_\sigma |\det \sigma(x)| \ ,$$

where the sum ranges over all $(m \times m)$-submatrices $\sigma(x)$ of $\nabla F(x)$. Note that $\nabla F(x)$ has rank m iff there exists at least one $(m \times m)$-submatrix $\sigma_0(x)$ of $\nabla F(x)$ satisfying $\det \sigma_0(x) \neq 0$.

openness part. Let $F \in \mathcal{P}_r$, i.e.,

$$\sigma_F(x) > 0 \quad \text{ for all } x \in \mathbb{R}^n \ .$$

Take any $\overline{x} \in \mathbb{R}^n$ (arbitrarily). Then

$$\eta_{\overline{x}} := \min_{x \in \text{cl } B(\overline{x}, 1)} \sigma_F(x) > 0 \ ,$$

and there exists $\varepsilon_{\overline{x}} > 0$ such that for any $G \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ satisfying

$$\|F - G\|_{k,\text{cl } B(\overline{x},1)} < \varepsilon_{\overline{x}}$$

(cf., (2.6)) we have

$$\min_{x \in \text{cl } B(\overline{x},1)} \sigma_G(x) > \frac{1}{2}\eta_{\overline{x}} \; . \tag{4.13}$$

We now can chose a sequence of points $x_i \in \mathbb{R}^n, \;\; i \in \mathbb{N}$, such that the open balls $\{B(x_i,1)\}_{i\in\mathbb{N}}$ cover $\mathbb{R}^n$ and that this cover is locally finite. Let further $\{\theta_i\}_{i\in\mathbb{N}}$ be a partition of unity subordinate to $\{B(x_i,1)\}_{i\in\mathbb{N}}$ (cf., Definition 4.1) and let $\varepsilon_{x_i}, \eta_{x_i}$ be defined as above: for all $i \in \mathbb{N}$ we have

$$\min_{x \in \text{cl } B(x_i,1)} \sigma_G(x) > \frac{1}{2}\eta_{x_i} \; \forall G \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \quad \text{with} \quad \|F - G\|_{k,\text{cl } B(x_i,1)} < \varepsilon_{x_i} \; .$$

Since $\{\text{supp}\,\theta_i\}_{i\in\mathbb{N}}$ is locally finite, for any $i_0 \in \mathbb{N}$ the set $I(i_0) = \{i \in \mathbb{N} \mid \text{supp}\,\theta_i \cap \text{supp}\,\theta_{i_0} \neq \emptyset\}$ is finite. We define $\overline{\varepsilon}_{i_0} = \min_{i\in I(i_0)} \varepsilon_{x_i}$ and the strictly positive continuous function

$$\rho(x) = \sum_{i\in\mathbb{N}} \overline{\varepsilon}_i \theta_i(x) \; .$$

For an arbitrary point $x_0 \in \text{supp}\,\theta_{i_0}$ we find

$$\rho(x_0) = \sum_{i\in I(i_0)} \overline{\varepsilon}_i \theta_i(x_0) \leq \sum_{i\in I(i_0)} \varepsilon_{x_{i_0}} \theta_i(x_0) \leq \varepsilon_{x_{i_0}} \; .$$

Consequently, by construction ($\text{supp}\,\theta_i \subset B(x_i,1)$), function $\rho(x)$ defines a C^k_s-neighborhood $U^k_{\rho,F}$ such that for any $G \in U^k_{\rho,F}$, the function $\sigma_G(x)$ is positive on $\mathbb{R}^n$ and thus G is contained in $\mathcal{P}_r$.

density part. Let $G \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ be an arbitrary function and let $U^k_{\phi,G}$ be an arbitrary C^k_s-neigborhood defined by the function $\phi \in C^0_+(\mathbb{R}^n, \mathbb{R})$.
We now construct a sequence of functions $F_i \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \cap U^k_{\phi,G}$, $i \in \mathbb{N}$, such that (the pointwise limit)

$$F := \lim_{i\to\infty} F_i \quad \text{ is contained in } \mathcal{P}_r \cap U^k_{\phi,G} \; .$$

Recall the open cover $\{B(x_i,1)\}_{i\in\mathbb{N}}$ of $\mathbb{R}^n$, and the corresponding partition of unity $\{\theta_i\}_{i\in\mathbb{N}}$ in the openness part. By Ex. 4.1 also $\{\text{int}\,\text{supp}\,\theta_i\}_{i\in\mathbb{N}}$ is an open cover of $\mathbb{R}^n$ and we can chose an partition of unity $\{\chi_i\}_{i\in\mathbb{N}}$ subordinate to $\{\text{int}\,\text{supp}\,\theta_i\}_{i\in\mathbb{N}}$. Note that the sets $\text{supp}\,\theta_i$ are closed and satisfy

$$\text{int}\,\text{supp}\,\theta_i \subset \text{supp}\,\theta_i \subset B(x_i,1) \; .$$

Thus we can find functions $\xi_i \in C^\infty(\mathbb{R}^n, \mathbb{R})$ satisfying (see Lemma 4.1)

$$0 \leq \xi_i(x) \leq 1 \; \forall x \in \mathbb{R}^n, \qquad \xi_i(x) = 1 \text{ on an open neighborhood } W_i \text{ of } \text{supp}\,\chi_i,$$
$$\text{supp}\,\xi_i \subset \text{int}\,\text{supp}\,\theta_i \; .$$

Starting with $i = 1$, according to Lemma 4.4 we can chose $u_1 \in \mathbb{R}^m$ (small enough) such that $H_1(x) := G(x) + u_1$ is contained in $\mathcal{P}_r$ and

$$\|\xi_1(x)u_1\|_{k,\text{supp}\,\theta_1} < \varepsilon_1 =: \min_{x\in\text{supp}\,\theta_1} \phi(x) \ . \tag{4.14}$$

By defining

$$F_1(x) := \begin{cases} G(x) + \xi_1(x)u_1 & x \in B(x_1, 1) \\ G(x) & x \in \mathbb{R}^n \setminus B(x_1, 1) \end{cases}$$

we have given a function $F_1 \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \cap U^k_{\phi,G}$ and F_1 is contained in $\mathcal{P}_r(\text{supp}\,\chi_1)$, where for given closed $C \subset \mathbb{R}^n$ we again put

$$\mathcal{P}_r(C) := \{F \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \mid \text{rank } \nabla F(x) \text{ has rank } m \ \ \forall x \in C \text{ with } F(x) = 0\} \ .$$

In the same way we define $F_2 \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \cap U^k_{\phi,G} \cap \mathcal{P}_r(\cup_{i=1}^2 \text{supp}\,\chi_i)$ by

$$F_2(x) := \begin{cases} F_1(x) + \xi_2(x)u_2 & x \in B(x_2, 1) \\ F_1(x) & x \in \mathbb{R}^n \setminus B(x_2, 1) \end{cases} ,$$

where $u_2 \in \mathbb{R}^m$ is chosen such that $F_2 \in \mathcal{P}_r(\text{supp}\,\chi_2)$ holds as well as

$$\|\xi_2(x)u_2\|_{k,\text{supp}\,\theta_2} < \varepsilon_2 =: \min_{x\in\text{supp}\,\theta_2} \phi(x) \ ,$$

and in case $\text{supp}\,\theta_1 \cap \text{supp}\,\theta_2 \neq \emptyset$ such that we have

$$F_2(x) \in \mathcal{P}_r(C) \quad \text{also for } C = \text{supp}\,\chi_1 \ .$$

With regard to the proof of the openness part of $\mathcal{P}_r$ the latter is possible by chosing u_2 in the part $\xi_2(x)u_2$ of F_2 small enough.
In this way after s steps we have constructed a function $F_s \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \cap U^k_{\phi,G}$,

$$F_s(x) := \begin{cases} F_{s-1}(x) + \xi_s(x)u_s & x \in B(x_s, 1) \\ f_{s-1}(x) & x \in \mathbb{R}^n \setminus B(x_s, 1) \end{cases} ,$$

satisfying $F_s \in \mathcal{P}_r(\text{supp}\,\chi_s)$ and by openness arguments also

$$F_s \in \mathcal{P}_r(\cup_{i=1}^{s-1} \text{supp}\,\chi_i) \ .$$

Since the cover $\{\text{supp}\,\theta_i\}_{i\in\mathbb{N}}$ is locally finite, for any $\overline{x} \in \mathbb{R}^n$ there exists a number $s(\overline{x})$ such that

$$F_{j_1}(\overline{x}) = F_{j_2}(\overline{x}) \quad \text{ for all } j_1, j_2 > s(\overline{x}) \ .$$

Thus for any $x \in \mathbb{R}^n$ the pointwise limit

$$F(x) := \lim_{i\to\infty} F_i(x)$$

is well defined and by construction $F \in C^\infty(\mathbb{R}^n, \mathbb{R}) \cap U^k_{\phi,G}$. Moreover, since $\{\text{supp}\,\chi_i\}_{i\in\mathbb{N}}$ constitutes a cover of $\mathbb{R}^n$, we also have $F \in \mathcal{P}_r(\mathbb{R}^n) = \mathcal{P}_r$.

□

An implication of Theorem 4.6 is that generically the Newton method for solving a system

$$F(x) = 0 \ , \qquad \text{where} \quad F \in C^\infty(\mathbb{R}^n, \mathbb{R}^n) \ , \tag{4.15}$$

works well. More precisely, we consider the Newton procedure: chose a starting point $x_0 \in \mathbb{R}^n$ and iterate

$$x_{k+1} = x_k - [\nabla F(x_k)]^{-1} F(x_k) \ , \qquad k = 0, 1, \dots \ . \tag{4.16}$$

The well-known local convergence result for this Newton method is as follows.

THEOREM 4.7. *Given $F \in C^2(\mathbb{R}^n, \mathbb{R}^n)$, let $F(\overline{x}) = 0$ for some $\overline{x} \in \mathbb{R}^n$ and suppose that the Jacobian*

$$\nabla F(\overline{x}) \quad \textit{is nonsingular} \ .$$

Then there exists an open neighborhood $U_{\overline{x}}$ of $\overline{x}$ such that starting with any $x_0 \in U_{\overline{x}}$ the Newton iteration (4.16) *converges quadratically to $\overline{x}$, i.e., with some $c > 0$ we have*

$$\|x_{k+1} - \overline{x}\| \le c \, \|x_k - \overline{x}\|^2 \ , \quad k = 0, 1, \dots \ .$$

According to this theorem, with respect to the Newton method (4.16), the nice function set in $C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ is given by

$$\mathcal{P}_r = \big\{ F \in C^\infty(\mathbb{R}^n, \mathbb{R}^n) \mid \quad \text{at any } \overline{x} \in \mathbb{R}^n \text{ with } F(\overline{x}) = 0 \text{ the Jacobian } \nabla F(\overline{x}) \text{ is nonsingular} \big\} \ . \tag{4.17}$$

As a special case of Theorem 4.6 we obtain the genericity result.

COROLLARY 4.1. [Genericity result for the Newton method]
The set $\mathcal{P}_r$ in (4.17) *is open and dense (thus generic) in $C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ wrt. the C^k_s-topology for any $k \ge 1$.*

4.5. Genericity results for constrained programs

In this section we perform a genericity analysis for *nonlinear programs* of the form (omitting equality constraints for notational simplicity),

$$(P): \qquad \min_x f(x) \qquad \text{s.t.} \qquad x \in \mathcal{F} := \{x \in \mathbb{R}^n \mid g_j(x) \le 0, \ j \in J\} \ , \tag{4.18}$$

with index set $J = \{1, \dots, m\}$ and *feasible set* $\mathcal{F}$. The functions f, g_j are assumed to be in $C^\infty(\mathbb{R}^n, \mathbb{R})$. To make explicit that (P) is given by the problem functions $f, G := (g_1, \dots, g_m)^T$ we often write $P(f, G)$, $\mathcal{F}(G)$ instead of (P), $\mathcal{F}$.

For $\overline{x} \in \mathcal{F}$ we introduce the *active index set* $J_{\overline{x}} = \{j \in J \mid g_j(\overline{x}) = 0\}$, and the *Lagrangean function* (near $\overline{x}$)

$$L(x, \mu) = f(x) + \sum_{j \in J_{\overline{x}}} \mu_j g_j(x) \ .$$

The coefficients μ_j are called *Lagrangean multipliers*.

DEFINITION 4.2. *We say that the* Linear Independence Constraint Qualification *(LICQ) is satisfied at $\overline{x} \in \mathcal{F}$ if*

$$\nabla g_j(\overline{x}), \ j \in J_{\overline{x}}, \ \textit{are linearly independent} \ .$$

We now state the well-known necessary optimality conditions.

LEMMA 4.5. [Necessary optimality conditions]
Let $\overline{x} \in \mathcal{F}$ be a local minimizer of (P) such that LICQ holds at $\overline{x}$. Then there exist multipliers $\overline{\mu}_j \geq 0,\ j \in J_{\overline{x}}$ (unique by LICQ), such that the Karush-Kuhn-Tucker (KKT) condition holds:

$$\nabla_x L(\overline{x}, \overline{\mu}) = \nabla f(\overline{x}) + \sum_{j \in J_{\overline{x}}} \overline{\mu}_j \nabla g_j(\overline{x}) = 0 \ . \tag{4.19}$$

Moreover the following second order conditions are satisfied:

$$d^T \nabla_x^2 L(\overline{x}, \overline{\mu}) d \geq 0 \quad \textit{for all } d \textit{ in } C_{\overline{x}} \ ,$$

where $C_{\overline{x}}$ is the cone of critical directions,

$$C_{\overline{x}} = \{d \in \mathbb{R}^n \mid \nabla f(\overline{x}) d \leq 0,\ \nabla g_j(\overline{x}) d \leq 0,\ j \in J_{\overline{x}}\} \ .$$

Proof. For a proof we refer to [13, Theorem 12.7].

□

A point $\overline{x} \in \mathcal{F}$ satisfying the KKT condition is called a KKT point.

EX. 4.2. *Let the KKT-condition (4.19) hold at $\overline{x} \in \mathcal{F}$. Then if LICQ is satisfied at $\overline{x}$, the multipliers $\overline{\mu}_j, j \in J_0(\overline{x})$, are uniquely determined.*

At a KKT point $\overline{x}$, with corresponding multipliers $\overline{\mu}_j,\ j \in J_{\overline{x}}$, strict complementarity is said to hold if

$$\overline{\mu}_j > 0, \ \ \forall j \in J_{\overline{x}} \ . \qquad \text{(SC)} \tag{4.20}$$

It is not difficult to show that if at a KKT point $\overline{x}$ the condition SC holds, then the cone $C_{\overline{x}}$ coincides with the tangent space $T_{\overline{x}}$ (see Ex. 4.3),

$$C_{\overline{x}} = T_{\overline{x}} := \{d \in \mathbb{R}^n \mid \nabla g_j(\overline{x}) d = 0,\ j \in J_{\overline{x}}\} \ . \tag{4.21}$$

EX. 4.3. *Let the KKT condition hold at $\overline{x}$ with $\overline{\mu}_j \geq 0,\ j \in J_{\overline{x}}$. Show that we have*

$$C_{\overline{x}} = \{d \mid \nabla g_j(\overline{x}) d = 0, \textit{ if } \overline{\mu}_j > 0, \nabla g_j(\overline{x}) d \leq 0, \textit{ if } \overline{\mu}_j = 0,\ j \in J_{\overline{x}}\} \ ,$$

i.e., $T_{\overline{x}} \subset C_{\overline{x}}$ for $T_{\overline{x}} := \{d \mid \nabla g_j(\overline{x}) d = 0,\ j \in J_{\overline{x}}\}$. If moreover SC holds, $C_{\overline{x}}$ equals the tangent space: $C_{\overline{x}} = T_{\overline{x}}$.

The necessary optimality conditions in Lemma 4.5 are not sufficient for optimality. We however obtain the following sufficiency and stability conditions.

THEOREM 4.8. *Let $P(f, G)$ be given for $(f, G) \in [C^2(\mathbb{R}^n, \mathbb{R})]^{1+m}$ and let $\overline{x} \in \mathcal{F}(G)$ satisfy LICQ, as well as the KKT condition* (4.19) *with SC and the* second order condition

$$(SOC): \qquad d^T \nabla_x^2 L(\overline{x}, \overline{\mu}) d > 0 \quad \forall d \in T_{\overline{x}} \setminus \{0\} \ .$$

(a) [Sufficient optimality conditions] *Then $\overline{x}$ is a strict local minimizer and an isolated KKT point of $P(f, G)$.*

(b) [Stability result] *There exist $\delta, \varepsilon > 0$, open neighborhoods $B(\overline{x}, \delta)$ of $\overline{x}$, and U_ε of (f, G) given by all $(\tilde{f}, \tilde{G}) \in [C^2(\mathbb{R}^n, \mathbb{R})]^{1+m}$ satisfying*

$$\|(\tilde{f}, \tilde{G}) - (f, G)\|_{2, \text{cl } B(\overline{x}, \delta)} < \varepsilon \ ,$$

as well as a (Frechet) differentiable map $x : U_\varepsilon \to B(\overline{x}, \delta)$ *with the following property: For all* $(\tilde{f}, \tilde{G}) \in U_\varepsilon$ *the point* $x(\tilde{f}, \tilde{G})$ *is a strict local minimizer and the unique KKT point of* $(P(\tilde{f}, \tilde{G}))$ *in* $\mathrm{cl}\, B(\overline{x}, \delta)$.

Proof. For (a) we refer to [13, Theorem 2.4,2.5]. The proof of the stability result in (b) depends on an application of the Implicit Function Theorem (in Banach spaces) to the system of KKT equations

$$H(x, \mu; f, G) := \begin{array}{rl} \nabla^T f(x) + \sum_{j \in J_{\overline{x}}} \mu_j \nabla^T g_j(x) & = 0 \\ g_j(x) & = 0, \quad j \in J_{\overline{x}} \end{array}, \tag{4.22}$$

(with $\mu_j \geq 0$ and x feasible) (see also [21, Theorem 6.1.5]). It is not difficult to show that under the assumptions of the theorem, the Hessian (recall $G_{J_{\overline{x}}} = (g_j, \ j \in J_{\overline{x}})^T$)

$$\nabla_{(x,\mu)} H(x, \mu; f, G) = \begin{pmatrix} \nabla_x^2 L(x, \mu) & \nabla G_{J_{\overline{x}}}(x)^T \\ \nabla G_{J_{\overline{x}}}(x) & 0 \end{pmatrix} \tag{4.23}$$

is nonsingular at $(\overline{x}, \overline{\mu})$ and at the given problem functions (f, G) (see, Ex.4.4).

□

EX. 4.4. *Let be given a symmetric matrix* $A \in S^n$ *and a matrix* $B \in \mathbb{R}^{n \times k}$ *with full rank* k*. Then for the matrix* $M := \left(\begin{smallmatrix} A & B \\ B^T & 0 \end{smallmatrix}\right)$ *the following holds.*
(a) $s^T A s \neq 0 \ \forall 0 \neq s \in \ker(B^T) \quad \Rightarrow \quad M$ *is nonsingular.*
One can even show $\Leftrightarrow$ *(see, e.g.,* [17, Lemma 2.43]).
(b) Suppose $s^T A s \geq 0 \ \ \forall s \in \ker(B^T)$ holds. Then

$$M \text{ is nonsingular} \quad \Leftrightarrow \quad s^T A s > 0 \ \forall 0 \neq s \in \ker(B^T).$$

Proof. The proof of (a) is straightforward. The statement (b) can be shown by using [21, Ex.3.2.22].

□

EX. 4.5. *Show that* $\overline{x} = (\frac{1}{\sqrt{2}}, \frac{1}{2})$ *is the unique (global) minimizer of*

$$(P): \qquad \min \ f(x) := x_1^2 + (x_2 - 1)^2 \quad \textit{s.t.} \quad x_2 \leq x_1^2, \ x_1 \geq 0 \ .$$

Hint: Show that $\overline{x}$ *satisfies the KKT conditions with LICQ, SC, and SOC.*

For later purposes we also need a stability result assuring that if no KKT point exists for $P(f, G)$ near a point $\overline{x}$, then after a sufficiently small perturbation of (f, G) this property remains true.

LEMMA 4.6. *Let* $(f, G) \in [C^\infty(\mathbb{R}^n, \mathbb{R})]^{1+m}$ *and* $\overline{x} \in \mathbb{R}^n$ *be given. Assume that there exists a neigborhood* $B(\overline{x}, \delta), \ \delta > 0$*, such that LICQ holds for all* $x \in \mathcal{F}(G) \cap \mathrm{cl}\, B(\overline{x}, \delta)$ *and that no KKT point of* $P(f, G)$ *is contained in* $\mathrm{cl}\, B(\overline{x}, \delta)$*. Then there exists* $\varepsilon > 0$ *and an open neigborhood* U_ε *of* (f, G) *defined by*

$$U_\varepsilon = \{(\tilde{f}, \tilde{G}) \in [C^\infty(\mathbb{R}^n, \mathbb{R})]^{1+m} \mid \|(\tilde{f}, \tilde{G}) - (f, G)\|_{1, \mathrm{cl}\, B(\overline{x}, \delta)} < \varepsilon\} \ ,$$

such that for all $(\tilde{f},\tilde{G}) \in U_\varepsilon$ *we have: the condition LICQ holds at all* $x \in \mathcal{F}(\tilde{G}) \cap \mathrm{cl}\, B(\overline{x},\delta)$ *and*

the set $\mathrm{cl}\, B(\overline{x},\delta)$ *does not contain any KKT point of* $P(\tilde{f},\tilde{G})$.

Proof. By a continuity argument (similar to the following reasoning) the statement about the stability of LICQ is easily proven. Let us now assume that the statement about the stability for KKT points is false. Then there must exist a sequence of functions $(\tilde{f}_\nu,\tilde{G}_\nu) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{1+m}$, $\nu \in \mathbb{N}$, satisfying

$$\|(\tilde{f}_\nu,\tilde{G}_\nu) - (f,G)\|_{1,\mathrm{cl}\, B(\overline{x},\delta)} \to 0 \;, \tag{4.24}$$

such that $P(\tilde{f}_\nu,\tilde{G}_\nu)$ possesses KKT points x_ν in $\mathrm{cl}\, B(\overline{x},\delta)$. The KKT conditions for x_ν read:

$$\begin{array}{rl} \nabla \tilde{f}_\nu(x_\nu) + \sum_{j\in J_{x_\nu}} \mu_j^\nu \nabla[\tilde{g}_j]_\nu(x_\nu) & = 0 \\ \left[\tilde{g}_j\right]_\nu(x_\nu) & = 0, \;\; j \in J_{x_\nu} \end{array} \qquad \nu \in \mathbb{N}\;. \tag{4.25}$$

Wlog. (by chosing a subsequence) we can assume (use (4.24))

$$x_\nu \to \tilde{x} \in \mathrm{cl}\,(B(\overline{x},\delta)\cap\mathcal{F}(G)),\;\; \nabla\tilde{f}_\nu(x_\nu) \to \nabla f(\tilde{x}),\;\; \nabla[g_j]_\nu(x_\nu) \to \nabla g_j(\tilde{x})\;. \tag{4.26}$$

Since there are only finitely many subsets $J_{x_\nu} \subset J$, we also can assume $J_{x_\nu} = J_0 \subset J_{\tilde{x}}$ for all ν ($J_{x_\nu} \subset J_{\tilde{x}}$ holds by continuity). We now show that the multipliers $\mu_j^\nu, j \in J_0$, are uniformly bounded by some constant M:

$$0 \le \mu_j^\nu \le M \qquad \forall j \in J_0,\; \forall \nu \in \mathbb{N}\;. \tag{4.27}$$

Let us suppose that this is not the case. Then we must have $\mu_0^\nu := \max_{j\in J_0} \mu_j^\nu \to \infty$ for $\nu \to \infty$. Wlog. we can assume $\mu_0^\nu = \mu_{j_0}^\nu$ for some $j_0 \in J_0$. By dividing the first relation in (4.25) by $\mu_{j_0}^\nu$ we find

$$\frac{\nabla\tilde{f}_\nu(x_\nu)}{\mu_{j_0}^\nu} + \nabla[\tilde{g}_{j_0}]_\nu(x_\nu) + \sum_{j\in J_0\setminus\{j_0\}} \frac{\mu_j^\nu}{\mu_{j_0}^\nu}\nabla[\tilde{g}_j]_\nu(x_\nu) = 0\;.$$

Letting $\nu \to \infty$, in view of $0 \le \frac{\mu_j^\nu}{\mu_{j_0}^\nu} \le 1$, we have (for some subsequence) $\frac{\mu_j^\nu}{\mu_{j_0}^\nu} \to \tilde{\mu}_j$ and thus (see (4.26))

$$\nabla g_{j_0}(\tilde{x}) + \sum_{j\in J_0\setminus\{j_0\}} \tilde{\mu}_j \nabla g_j(\tilde{x}) = 0\;,$$

contradiction the fact that LICQ holds at $\tilde{x} \in \mathrm{cl}\,(B(\overline{x},\delta)\cap\mathcal{F}(G))$. In view of (4.27) we conclude (for some subsequence of ν),

$$\mu_j^\nu \to \tilde{\mu}_j,\; j \in J_0, \;\; \text{for } \nu \to \infty\;,$$

and by taking in (4.25) the limit $\nu \to \infty$ we find

$$\begin{array}{rl} \nabla f(\tilde{x}) + \sum_{j\in J_0} \tilde{\mu}_j \nabla g_j(\tilde{x}) & = 0 \\ g_j(\tilde{x}) & = 0, \;\; j \in J_0 \end{array} \;.$$

Thus, recalling $J_0 \subset J_{\tilde{x}}$, $\tilde{x}$ is a KKT point of $P(f,G)$ in $\mathrm{cl}\, B(\overline{x},\delta)$ contradicting our assumption.

□

An important type of solution method for computing local minimizers of $P(f,G)$ are so-called Newton methods. These methods apply (some sort of) Newton iteration to the system (4.22) of KKT equations for $P(f,G)$.
In the proof of Theorem 4.8(b) we have seen that if a KKT point $\overline{x}$ satisfies LICQ, SC, and SOC, then the Jacobian (wrt. (x,μ), see (4.23)) of the system (4.22) is nonsingular, and according to Theorem 4.7 the Newton iteration applied to the KKT equation (4.22) is locally quadratically convergent to $\overline{x}$. So, a desirable property is that the problem functions (f,G) of $P(f,G)$ belong to the following "nice set"

$$\text{(4.28)} \quad \mathcal{P}_r := \{(f,G) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{1+m} \mid \text{at all local minimizers } \overline{x} \text{ of } P(f,G) \text{ we have LICQ, SC, and SOC}\} \; .$$

Here, $n,m \in \mathbb{N}$ are given numbers defining the "size" of problem $P(f,G)$. A local minimizer satisfying LICQ, SC, and SOC, is called a nondegenerate minimizer.

We will now prove that this set $\mathcal{P}_r$ is open and dense in the C^k_s-topology for any $k \geq 2$. We need two auxiliary lemmas. As a corollary of Theorem 4.6 we now show that generically the LICQ condition is valid at all feasible points in $\mathcal{F}(G)$.

LEMMA 4.7.
(a) Let $G = (g_1,\dots,g_m)^T \in C^\infty(\mathbb{R}^n,\mathbb{R}^m)$ be given. Then for almost all $a \in \mathbb{R}^m$ we have: for the perturbed function $\tilde{G}(x) = G(x) + a$ at all feasible points in $\mathcal{F}(\tilde{G})$ the condition LICQ is valid.
(b) The subset of constraint functions $G \in C^\infty(\mathbb{R}^n,\mathbb{R}^m)$ such that at all feasible points in $\mathcal{F}(G)$ the condition LICQ holds is open and dense in $C^\infty(\mathbb{R}^n,\mathbb{R}^m)$ wrt. the C^k_s-topology for all $k \geq 1$.

Proof. Chose any subset $J_0 \subset J = \{1,\dots,m\}$ and consider the feasible points $x \in \mathcal{F}(G)$ with $J_x = J_0$, i.e, x solves the system

$$G_{J_0}(x) := \begin{pmatrix} g_j(x) \\ \vdots \end{pmatrix}_{j \in J_0} = 0 \; ,$$

of $|J_0|$ equations in n unknowns.
Note that LICQ at x is equivalent with the condition that $\nabla G_{J_0}(x)$ has rank $|J_0|$. By Theorem 4.6 the set of functions $F(x) := G_{J_0}(x) \in C^\infty(\mathbb{R}^n,\mathbb{R}^{|J_0|})$ such that LICQ holds for all x satisfying $G_{J_0}(x) = 0$, is open and dense in the C^k_s-topology for any $k \geq 1$. By noticing that there only exist finitely many subsets J_0 of J, by taking the intersection of the corresponding sets we have proven the statement.

□

In the next lemma, for given problem functions $(f,G) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{1+m}$, we consider the perturbations $(\tilde{f},\tilde{G}),\ \tilde{G} = (\tilde{g}_1,\dots,\tilde{g}_m)^T$ defined by

$$\text{(4.29)} \qquad \begin{array}{ll} \tilde{f}(x) & = f(x) + b^T x \\ \tilde{g}_j(x) & = g_j(x) + \alpha_j \; , \quad j \in J \end{array} \; ,$$

depending on the parameter $(b,a) = (b,\alpha_1,\dots,\alpha_m) \in \mathbb{R}^{n+m}$.

LEMMA 4.8. *Let $(f, G) \in [C^\infty(\mathbb{R}^n, \mathbb{R})]^{1+m}$ be given. Then for almost all parameters $(b, a) \in \mathbb{R}^{n+m}$ the perturbed function $(\tilde{f}, \tilde{G})$ in (4.29) belongs to $\mathcal{P}_r$.*

Proof. In view of Lemma 4.7(a) for almost all $a \in \mathbb{R}^m$ the set $\mathcal{F}(\tilde{G})$ satisfies LICQ. Thus in view of Lemma 4.5 for these a all local minimizers must be KKT points. So we can consider KKT solutions $(\overline{x}, \overline{\mu})$ of $P(\tilde{f}, \tilde{G})$ with a given active index set $J_0 = J_{\overline{x}}$, i.e., $(\overline{x}, \overline{\mu})$, $\overline{\mu} \in \mathbb{R}_+^{|J_0|}$ solve the KKT equations

$$(4.30) \qquad H(x, \mu; \tilde{f}, \tilde{G}) := \begin{array}{rl} \nabla^T \tilde{f}(x) + \sum_{j \in J_0} \mu_j \nabla^T \tilde{g}_j(x) & = 0 \\ \tilde{g}_j(x) & = 0, \quad j \in J_0 \end{array} .$$

We first show that for almost all $(b, a) \in \mathbb{R}^{n+m}$ at all KKT points of $P(\tilde{f}, \tilde{G})$ the SC condition is met. Suppose that at a KKT point $(\overline{x}, \overline{\mu})$, i.e., a solution of (4.30), SC does not hold. Wlog. we assume $J_0 = \{1, \ldots, s\}$. This means that for at least one $j \in J_0$ ($J_0 \neq \emptyset$) say $j = 1$, the condition $\overline{\mu}_1 = 0$ is satisfied. Thus the KKT point $(\overline{x}, \overline{\mu})$ is a solution of the system

$$(4.31) \qquad \begin{array}{rl} H(x, \mu; \tilde{f}, \tilde{G}) = 0 & \text{see (4.30) with } J_{\overline{x}} = J_0 \\ \mu_1 = 0 & \end{array} .$$

The Jacobian of this system wrt. (x, μ, b, a) has the form (e_1 is the unit vector $(1, 0, .., 0) \in \mathbb{R}^{|J_0|}$)

$$\begin{pmatrix} \nabla_x^2 L(x, \mu) & \nabla G_{J_0}(x)^T & I_n & 0 \\ \nabla G_{J_0}(x) & 0 & 0 & I_{|J_0|} \\ 0 & e_1^T & 0 & 0 \end{pmatrix} .$$

By adding an appropriate multiple of the last row to the first n rows the

$$\text{submatrix} \quad \begin{array}{c} \nabla G_{J_0}(x)^T \\ 0 \\ e_1^T \end{array} \quad \text{is transformed to} \quad \begin{array}{cccc} 0 & \nabla^T g_2(x) & \ldots & \nabla^T g_s(x) \\ 0 & 0 & \ldots & 0 \\ 1 & 0 & \ldots & 0 \end{array} ,$$

while the other parts remain unchanged. Obviously the obtained matrix has full row rank $n + |J_0| + 1$. According to the Parametric Sard Theorem, for almost all (b, a) also the Jacobian of the system (4.31) wrt. the $n + |J_0|$ variables (x, μ) has full row rank $n + |J_0| + 1$ at all solutions $(\overline{x}, \overline{\mu})$ of (4.31). Consequently for almost all (b, a) solutions (x, μ) of (4.31) (with $\mu_1 = 0$) are excluded.
Now we check the SOC condition (assuming that SC holds). To do so, we look at the Jacobian of $H(x, \mu; \tilde{f}, \tilde{G})$ wrt. to (x, μ, b, a), which has the form (recall the abbreviation $G_{J_0}(x) = (g_j(x), j \in J_0)^T$)

$$\nabla_{(x,\mu,b,a)} H(x, \mu; \tilde{f}, \tilde{G}) = \begin{pmatrix} \nabla_x^2 L(x, \mu) & \nabla G_{J_0}(x)^T & I_n & 0 \\ \nabla G_{J_0}(x) & 0 & 0 & I_{|J_0|} \end{pmatrix} .$$

Obviously, this matrix has full row rank $n + |J_0|$ and the Parametric Sard Theorem 4.2 implies that for almost all $(b, a) \in \mathbb{R}^{n+m}$ the following is true: at all solutions $(\overline{x}, \overline{\mu})$ of (4.30) the Jacobian

$$(4.32) \qquad \nabla_{(x,\mu)} H(\overline{x}, \overline{\mu}; \tilde{f}, \tilde{G}) = \begin{pmatrix} \nabla_x^2 L(\overline{x}, \overline{\mu}) & \nabla G_{J_0}(\overline{x})^T \\ \nabla G_{J_0}(\overline{x}) & 0 \end{pmatrix}$$

has full rank $n+|J_0|$. Note that by the result of Lemma 4.4 (and the observation before this lemma), for almost all $a \in \mathbb{R}^{|J_0|}$ the case $|J_0| > n$ is excluded. So we can assume $|J_0| \leq n$ and that the matrix $\nabla G_{J_0}(x)$ has full rank $|J_0|$ at all solutions $(\overline{x}, \overline{\mu})$ of (4.30) (due to LICQ, cf., Lemma 4.7(a)). With regard to the necessary (second order) optimality conditions in Lemma 4.5, by applying Ex 4.4 (use $T_{\overline{x}} = C_{\overline{x}}$ because of SC) we conclude that for almost all $(b, a) \in \mathbb{R}^{n+m}$ at all minimizers $\overline{x}$ of $P(\tilde{f}, \tilde{G})$ with $J_{\overline{x}} = J_0$, the condition LICQ, SC and SOC must hold.
Since there only exist finitely many subsets J_0 of J, by taking the corresponding intersections of the perturbations (b, a), the statement of the lemma is shown.

□

We finally can formulate our main genericity result.

THEOREM 4.9. [Genericity result for constrained programs]
The set $\mathcal{P}_r$ in (4.28) *is open and dense in $[C^\infty(\mathbb{R}^n, \mathbb{R})]^{1+m}$ wrt. the C^k_s-toloplogy for any $k \geq 2$.*

Proof. To avoid the technicalities involved with a function σ which characterizes the adherence to $\mathcal{P}_r$ as in the proof of [21, Th. 7.1.13] (see the functions σ_f, σ_F in the proofs of Theorem 4.5, 4.6) we base our proof on the stability result in Theorem 4.8, and on Lemmas 4.6,4.7.

openness part. Let $(f, G) \in \mathcal{P}_r$. Then $P(f, G)$ has a sequence $\overline{x}_i,\ i \in I \subset \mathbb{N}$ (I possibly infinite) of local minimizers satisfying the conditions of Theorem 4.8 with corresponding $\varepsilon_{\overline{x}_i}, B(\overline{x}_i, \delta_{\overline{x}_i})$ and has no further KKT points. According to Lemma 4.7(b) we can assume that LICQ holds for $\mathcal{F}(G)$. So we can select an infinite sequence of points $\hat{x}_i,\ i \in \hat{I} \subset \mathbb{N}$, such that with corresponding $\varepsilon_{\hat{x}_i}, B(\hat{x}_i, \delta_{\hat{x}_i})$ the conditions in Lemma 4.6 are satisfied and such that the sets $\{B(\overline{x}_i, \delta_{\overline{x}_i})\}_{i \in I} \cup \{B(\hat{x}_i, \delta_{\hat{x}_i})\}_{i \in \hat{I}}$ form a locally finite open cover of $\mathbb{R}^n$. We re-arrange the points $\hat{x}_i, \overline{x}_i$ into one sequence

$$\{x_i\}_{i \in \mathbb{N}} = \{\overline{x}_i\}_{i \in I} \cup \{\hat{x}_i\}_{i \in \hat{I}} \quad \text{with corresponding} \quad \varepsilon_{x_i}, B(x_i, \delta_{x_i}) \ .$$

By construction, for any $i \in \mathbb{N}$ it holds: For any $(\tilde{f}, \tilde{G}) \in [C^\infty(\mathbb{R}^n, \mathbb{R})]^{m+1}$ satisfying

$$\|(\tilde{f}, \tilde{G}) - (f, G)\|_{k, \text{cl}\, B(x_i, \delta_{x_i})} < \varepsilon_{x_i} \ ,$$

the problem $P(\tilde{f}, \tilde{G})$ either has a unique local minimizer (unique KKT point) in $\text{cl}\, B(x_i, \delta_{x_i})$ satisfying the conditions of Theorem 4.8, or has no KKT points in $\text{cl}\, B(x_i, \delta_{x_i})$. We now take a partition of unity $\{\theta_i\}_{i \in \mathbb{N}}$ subordinate to $\{B(x_i, \delta_{x_i})\}_{i \in \mathbb{N}}$ and define for any $i_0 \in \mathbb{N}$ the finite set $I(i_0) = \{i \in \mathbb{N} \mid \text{supp}\, \theta_i \cap \text{supp}\, \theta_{i_0} \neq \emptyset\}$ (finite, since the cover $\{B(x_i, \delta_{x_i})\}_{i \in \mathbb{N}}$ is locally finite). As before we put $\overline{\varepsilon}_{i_0} := \min_{i \in I(i_0)} \varepsilon_{x_i}$ and introduce the strictly positive continuous function

$$\rho(x) = \sum_{i \in \mathbb{N}} \overline{\varepsilon}_i \theta_i(x) \ .$$

For an arbitrary point $x_0 \in \text{supp}\, \theta_{i_0}$ we find

$$\rho(x_0) = \sum_{i \in I(i_0)} \overline{\varepsilon}_i \theta_i(x_0) \leq \sum_{i \in I(i_0)} \varepsilon_{x_{i_0}} \theta_i(x_0) \leq \varepsilon_{x_{i_0}} \ .$$

Consequently, by construction, the function $\rho(x)$ defines a C_s^k-neighborhood $U^k_{\rho,(f,G)}$ such that any $(\tilde{f},\tilde{G}) \in U^k_{\rho,(f,G)}$ is contained in $\mathcal{P}_r$.

density part. We shall argue as in the proofs of Theorems 4.5, 4.6. Let $(\hat{f},\hat{G}) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{m+1}$ be an arbitrary (fixed) pair of problem functions and let $U^k_{\phi,(\hat{f},\hat{G})}$ be an arbitrary C_s^k-neigborhood defined by $\phi \in C^0_+(\mathbb{R}^n,\mathbb{R})$ (see Definition 2.2). We again construct a sequence of functions $(f_i,G_i) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{m+1} \cap U^k_{\phi,(\hat{f},\hat{G})}$, $i \in \mathbb{N}$, such that (the pointwise limit)

$$(f,G) := \lim_{i\to\infty}(f_i,G_i) \quad \text{is contained in } \mathcal{P}_r \cap U^k_{\phi,(\hat{f},\hat{G})} \ .$$

To that end chose points $x_i \in \mathbb{N}$, such that $\{B(x_i,1)\}_{i\in\mathbb{N}}$ constitutes a locally finite cover of $\mathbb{R}^n$, and take a corresponding partition of unity $\{\theta_i\}_{i\in\mathbb{N}}$. Recall that by Ex. 4.1 also $\{\text{int supp}\,\theta_i\}_{i\in\mathbb{N}}$ is an open cover of $\mathbb{R}^n$ and we can chose a partition of unity $\{\chi_i\}_{i\in\mathbb{N}}$ subordinate to $\{\text{int supp}\,\theta_i\}_{i\in\mathbb{N}}$. In view of

$$\text{int supp}\,\theta_i \subset \text{supp}\,\theta_i \subset B(x_i,1) \ ,$$

we can find functions $\xi_i \in C^\infty(\mathbb{R}^n,\mathbb{R})$ satisfying (see Lemma 4.1)

$$0 \le \xi_i(x) \le 1 \ \forall x \in \mathbb{R}^n \ , \qquad \begin{array}{l} \xi_i(x) = 1 \text{ on an open neighborhood } W_i \text{ of supp}\,\chi_i \ , \\ \text{supp}\,\xi_i \subset \text{int supp}\,\theta_i \ . \end{array}$$

Starting with $i=1$, according to Lemma 4.8 we can chose $(b_1,a_1) \in \mathbb{R}^{n+m}$ (small enough) such that $\big(\hat{f}(x) + b_1^T x, \hat{G}(x)) + a_1\big)$ is contained in $\mathcal{P}_r$ and

$$\big\|\xi_1(x)\big(b_1^T x, a_1\big)\big\|_{k,\text{supp}\,\theta_1} < \varepsilon_1 =: \min_{x\in\text{supp}\,\theta_1} \phi(x) \ . \tag{4.33}$$

By defining

$$\big(f_1(x),G_1(x)\big) := \left\{ \begin{array}{ll} \big(\hat{f}(x) + \xi_1(x) b_1^T x, \hat{G}(x) + \xi_1(x) a_1\big) & x \in B(x_1,1) \\ \big(\hat{f}(x), \hat{G}(x)\big) & x \in \mathbb{R}^n \setminus B(x_1,1) \end{array} \right. ,$$

we have given a function $(f_1,G_1) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{m+1} \cap U^k_{\phi,(\hat{f},\hat{G})}$, which also is contained in $\mathcal{P}_r(\text{supp}\,\chi_1)$, where for given closed $C \subset \mathbb{R}^n$ we put

$$\begin{array}{rl} \mathcal{P}_r(C) := \{(f,G) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{m+1} \mid & \text{all local minimizers } x \in C \\ & \text{of } P(f,G) \text{ are nondegenerate}\} \ . \end{array}$$

In the same way we define $(f_2,G_2) \in [C^\infty(\mathbb{R}^n,\mathbb{R})]^{m+1} \cap U^k_{\phi,(\hat{f},\hat{G})} \cap \mathcal{P}_r(\cup_{i=1}^2 \text{supp}\,\chi_i)$ by

$$\big(f_2(x),G_2(x)\big) := \left\{ \begin{array}{ll} \big(f_1(x) + \xi_2(x) b_2^T x, G_2(x) + \xi_2(x) a_2\big) & x \in B(x_2,1) \\ \big(f_1(x), G_1(x)\big) & x \in \mathbb{R}^n \setminus B(x_2,1) \end{array} \right. ,$$

where $(b_2,a_2) \in \mathbb{R}^{n+m}$ is chosen such that

$$\|\big(\xi_2(x) b_2^T x, \xi_2(x) a_2\big)\|_{k,\text{supp}\,\theta_2} < \varepsilon_2 =: \min_{x\in\text{supp}\,\theta_2} \phi(x) \ ,$$

and thus $(f_2, G_2) \in \mathcal{P}_r(\text{supp}\, \chi_2)$ holds, and such that in case $\text{supp}\, \theta_1 \cap \text{supp}\, \theta_2 \neq \emptyset$, we have

$$(f_2, G_2) \in \mathcal{P}_r(C) \quad \text{also for } C = \text{supp}\, \chi_1 \ .$$

With regard to the proof of the openness part of $\mathcal{P}_r$ the latter is possible by chosing the vectors b_2, a_2 of (f_2, G_2) small enough.

In this way after s steps we have constructed a function $(f_s, G_s) \in [C^\infty(\mathbb{R}^n, \mathbb{R})]^{m+1} \cap U^k_{\phi,(\hat{f},\hat{G})}$,

$$\big(f_s(x), G_s(x)\big) := \begin{cases} \big(f_{s-1}(x) + \xi_s(x) b_s^T x, G_{s-1}(x) + \xi_s(x) a_s\big) & x \in B(x_s, 1) \\ \big(f_{s-1}(x), G_{s-1}(x)\big) & x \in \mathbb{R}^n \setminus B(x_s, 1) \end{cases} ,$$

satisfying $(f_s, G_s) \in \mathcal{P}_r(\text{supp}\, \chi_s)$ and by openness arguments also

$$(f_s, G_s) \in \mathcal{P}_r(\cup_{i=1}^{s-1} \text{supp}\, \chi_i) \ .$$

Since the cover $\{\text{supp}\, \theta_i\}_{i \in \mathbb{N}}$ is locally finite, for any $\overline{x} \in \mathbb{R}^n$ there exists a number $s(\overline{x})$ such that

$$\big(f_{j_1}(\overline{x}), G_{j_1}(\overline{x})\big) = \big(f_{j_2}(\overline{x}), G_{j_2}(\overline{x})\big) \quad \text{for all } j_1, j_2 > s(\overline{x}) \ .$$

Thus the pointwise limit

$$\big(f(x), G(x)\big) := \lim_{i \to \infty} \big(f_i(x), G_i(x)\big) \ , \quad x \in \mathbb{R}^n \ ,$$

is well defined and by construction $(f, G) \in [C^\infty(\mathbb{R}^n, \mathbb{R})]^{m+1} \cap U^k_{\phi,(\hat{f},\hat{G})}$. Moreover, since $\{\text{supp}\, \chi_i\}_{i \in \mathbb{N}}$ constitutes a cover of $\mathbb{R}^n$, we also have $(f, G) \in \mathcal{P}_r(\mathbb{R}^n) = \mathcal{P}_r$.

□

CHAPTER 5

Manifolds, stratifications, and transversality

The goal of this chapter is to provide the basic techniques and results needed to formulate and to prove the genericity results for parametric problems in the next chapter.

5.1. Manifolds in $\mathbb{R}^n$

Manifolds in $\mathbb{R}^n$ of dimension d are subsets of $\mathbb{R}^n$ which locally are defined as solution sets of n-d independent equations.

DEFINITION 5.1. *Let $k \in \mathbb{N} \cup \{\infty\},\ k \geq 1$. A subset $M \subset \mathbb{R}^n$ is called a C^k-manifold of dimension $d = n - r$ ($0 \leq d \leq n$), or of codimension r ($0 \leq r \leq n$) (notation* dim $M = d$ *or codim $M = r$), if the following condition CM1 is satisfied.*

<u>CM1</u>: *To any $\overline{x} \in M$ there exist a $\mathbb{R}^n$-neighborhood U of $\overline{x}$ and r functions $f_j : U \to \mathbb{R}$, $f_j \in C^k(U, \mathbb{R}), j = 1, \cdots, r$, such that*

(i) $M \cap U = \{x \in U \mid f_j(x) = 0,\ j = 1, \cdots, r\}$,

(ii) $\nabla f_j(\overline{x}),\ j = 1, \ldots, r$, *are linearly independent.*

Let $M_1,\ M$ be manifolds in $\mathbb{R}^n$ of codimensions, $r_1,\ r$, respectively. Then if $M_1 \subset M$ holds, we say that M_1 is a submanifold of M of codimension $r_1 - r$ in M.

Manifolds in $\mathbb{R}^n$ of dimension d can also be seen as subsets which locally look like pieces of the Cartesian d-space $\mathbb{R}^d$. This is a consequence of the following equivalent condition.

<u>CM2</u>: Let be given $M \subset \mathbb{R}^n,\ d \in \mathbb{N}, 0 \leq d \leq n$, such that to any $\overline{x} \in M$ there exist $\mathbb{R}^n$-neighborhoods U of $\overline{x}, V$ of $0 \in \mathbb{R}^n$, and a map $\Phi : U \to V$ satisfying with $r = n - d$:

(i) $\Phi(U \cap M) = V \cap (0^r \times \mathbb{R}^d)$

(ii) Φ is a (local) C^k-diffeomorphism ($k \geq 1$) .

Here $0^r \times \mathbb{R}^d = \{y \in \mathbb{R}^n \mid y_1 = y_2 = \cdots = y_r = 0\}$.

THEOREM 5.1. *Let be given $M \subset \mathbb{R}^n, d \in \mathbb{N}, 0 \leq d \leq n,\ r = n - d$, and $k \in \mathbb{N} \cup \{\infty\}, k \geq 1$. Then, the conditions $CM1$ and $CM2$ are equivalent.*

Proof. "$CM1 \Rightarrow CM2$": Let $\overline{x} \in M, U, f_j, j = 1, \ldots, r$, as in $CM1$. By $CM1$ (ii) we can choose $n{-}r$ vectors $\xi_{r+1}, \ldots, \xi_n \in \mathbb{R}^n$, such that $\nabla^T f_1(\overline{x}), \ldots, \nabla^T f_r(\overline{x}), \xi_{r+1}, \ldots, \xi_n$, form a basis of $\mathbb{R}^n$. Putting

$$\Phi(x) = \big(f_1(x), \ldots, f_r(x), \xi_{r+1}^T(x - \overline{x}), \ldots, \xi_n^T(x - \overline{x})\big)^T$$

it follows $\Phi(\overline{x}) = 0$ and $\nabla\Phi(\overline{x})$ is nonsingular. By the Inverse Function Theorem 7.1 there exist a $\mathbb{R}^n$-neighbourhood $\hat{U}$ of $\overline{x}$, $\hat{U} \subset U$ and a $\mathbb{R}^n$-neighbourhood V of 0 such that $\Phi : \hat{U} \to V$ is a C^k-diffeomorphism (see Remark 7.1). Moreover by $CM1$ (i) we find

$$\Phi(\hat{U} \cap M) = \{y = \Phi(x) \mid x \in \hat{U} \cap M\} = \{y \in V \mid y_j = f_j(x) = 0, j = 1, \dots, r\}.$$

"$CM2 \Rightarrow CM1$": Let $\overline{x} \in M, U, V, \Phi$ as in $CM2$. Putting $f_j = \Phi_j, j = 1, \dots, r$, it follows by $CM2$ that $U \cap M = \Phi^{-1}(V \cap (0^r \times \mathbb{R}^d)) = \{x \in U \mid f_j(x) = 0, j = 1, \dots, r\}$. By Remark 7.1 the matrix $\nabla\Phi(\overline{x})$ is nonsingular. Consequently, the rows $\nabla f_j(\overline{x}) = \nabla\Phi_j(\overline{x}), j = 1, \dots, r$, of $\nabla\Phi(\overline{x})$ are linearly independent.

□

REMARK 5.1. Commonly a subset M of $\mathbb{R}^n$ is called a d-dimensional C^k-manifold (submanifold of $\mathbb{R}^n$) if M is locally C^k-diffeomorph to an open subset of $\mathbb{R}^d$ in the following sense:

CM3: To any $\overline{x} \in M$ there exists an open subset $\tilde{U}$ of M containing $\overline{x}$, and a C^k-diffeomorphism $\tilde{\Phi} : \tilde{U} \to \tilde{V}$ with $\tilde{V}$ open in $\mathbb{R}^d$.

This condition CM3 is equivalent with the relations in CM1 and CM2. We refer to [21, Theorem 3.1.1] for a proof.
More generally, a C^k-manifold M is often introduced as a locally Euclidean, second countable Hausdorff space (not necessarily a subset of $\mathbb{R}^n$) such that CM3 holds (see, e.g., [39, Definition 5.2]).
In our definition, a manifold M is always a subset (submanifold) of $\mathbb{R}^n$ and as such M inherits from $\mathbb{R}^n$ the Hausdorff topology with a countable base (see [32, p.5, 1.2.1 Definition and following observation]

Note that by these equivalent conditions CM1, CM2, and CM3, we have defined smooth manifolds of smoothness class k for $k \geq 1$. We give some simple examples of manifolds.

EX. 5.1.
(a) Given $0 \leq r \leq n$, the hyperplane

$$H^{n,r} = \{x \in \mathbb{R}^n \mid x_1 = \ldots = x_r = 0\}$$

is a C^∞-manifold of codimension r and of dimension $n - r$. This is clear from Definition 5.1, taking $f_j(x) = x_j,\ j = 1, \dots, r$..
(b) The sphere S_{n-1} in $\mathbb{R}^n$,

$$S_{n-1} = \{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i^2 = 1\}$$

is a C^∞-manifold with codimension $r = 1$ and dimension $n - 1$. Here again, $CM1$ is satisfied. The relation $f_1(x) := \sum_{i=1}^n x_i^2 - 1 = 0$ defines S_{n-1} globally.
Note that $\nabla f_1(x) = 2(x_1, \dots, x_n) \neq 0$ for all $x \in S_{n-1}$, i.e., $CM1$ (ii) is valid.
(c) Any linear subspace S spanned by d given linearly independent vectors $\xi_1, \dots, \xi_d \in$

$\mathbb{R}^n$,

$$S = \{x = \sum_{j=1}^{d} \alpha_j \xi_j \ , \ \alpha_j \in \mathbb{R}\},$$

is a C^∞-manifold of dimension d, codimension $n-d$. In fact, by choosing $n-d$ linearly independent vectors $\xi_{d+1}, \ldots, \xi_n \in \mathbb{R}^n$ which are perpendicular to S (i.e. perpendicular to the basis vectors $\xi_1, \ldots, \xi_d$ of S), the set S is defined (globally) as the solution set of the $n-d$ equations

$$f_\ell(x) = \xi_\ell^T x = 0 \ , \ \ell = d+1, \ldots, n.$$

(d) Any open set $M_0 \subset \mathbb{R}^n$ is a C^∞-manifold of dimension n (codimension $r = 0$). Indeed, $CM1$ is valid with $r = 0$.

(e) Let M^2 denote the set of real 2×2 matrices,

$$M^2 = \left\{ A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \ a_{ij} \in \mathbb{R} \right\},$$

which can be identified with $\mathbb{R}^4$. Then, the subset

$$\mathcal{A} = \{A \in M^2 \mid \det A = 0\} \setminus \{0\}$$

is a C^∞-submanifold of M^2 with dimension 3, codimension 1. To see this, consider the equation

$$f_1(A) := detA = a_{11}a_{22} - a_{12}a_{21} = 0.$$

The gradient $\nabla f_1(A) = (a_{22}, -a_{21}, -a_{12}, a_{11})$ is not equal to 0 if $A \neq 0$, i.e., $\nabla f_1(A) \neq 0$ for all $A \in \mathcal{A}$. Thus $CM1$ is valid.

Let $M \subset \mathbb{R}^n$ be a d-dimensional manifold. By a tangent vector wrt. M at $\overline{x} \in M$ we mean a vector $\xi \in \mathbb{R}^n$, such that there is a C^1-curve through $\overline{x}$ in M, with tangent vector to the curve in $\overline{x}$ equal to ξ. Since M is of dimension d there are d linearly independent directions ξ with such a property, spanning a d-dimensional space, the tangent space $T_{\overline{x}}M$. We will give a definition of that space based on Definition 5.1.

DEFINITION 5.2. *Let $M \subset \mathbb{R}^n$ be a C^k-manifold $(k \geq 1)$ of dimension d (codimension $r = n - d$) as defined in CM1. Let $\overline{x} \in M$. The linear space*

$$N_{\overline{x}}M = span\{\nabla f_j(\overline{x}), \ j = 1, \ldots, r\}$$

is called the normal space of M at $\overline{x}$, and the space

$$T_{\overline{x}}M = \{\xi \in \mathbb{R}^n \mid \nabla f_j(\overline{x})\xi = 0, \ j = 1, \ldots, r\} \ ,$$

is called the tangent space of M at $\overline{x}$. By definition, $\dim T_{\overline{x}}M = d$, $\dim N_{\overline{x}}M = r$.

The following lemma shows that the normal- and the tangent spaces in Definition 5.2 are well-defined.

LEMMA 5.1. *The spaces $N_{\overline{x}}M$ and $T_{\overline{x}}M$ in Definition 5.2 do not depend on the (possibly different systems of) functions $f_j, \ j = 1, \ldots, r$, defining the manifold M around $\overline{x} \in M$.*

Proof. Let $\overline{x} \in M, U, f_j, j = 1, \dots, r$, be as in CM1. Consider the corresponding local diffeomorphism $\Phi : U \cap M \to V \cap (0^r \times \mathbb{R}^d)(d = n - r)$ (cf., $CM2$ and Theorem 5.1). Then, with the matrices $F = \big(\nabla^T f_1(\overline{x}), \dots, \nabla^T f_r(\overline{x})\big)$ and $X = (\xi_{r+1}, \dots, \xi_n)$ (cf., the proof of Theorem 5.1) we have $\nabla\Phi(\overline{x}) = \binom{F^T}{X^T}$. We can assume that $S := \{\xi_j,\ j = r+1, \dots, n\}$ forms an orthonormal system and that it is orthogonal to $N_{\overline{x}}M = \text{span}\,\{\nabla f_j(\overline{x}), j = 1, \dots, r\}$, i.e., span $S = T_{\overline{x}}M,\ F^T X = 0$, and $X^T X = I_d$. Suppose, we have given another system of functions $\tilde{f}_j,\ j = 1, \dots, r$, defining M locally in U. (By making U small enough we can assume that U is the same.) Let $\tilde{\Phi} : U \cap M \to V \cap (0^r \times \mathbb{R}^d)$ be the corresponding diffeomorphism. Then again, with $\widetilde{F} := \big(\nabla^T \tilde{f}_1(\overline{x}), \dots, \nabla^T \tilde{f}_r(\overline{x})\big)$ we can assume $\nabla\tilde{\Phi}(\overline{x}) = \binom{\widetilde{F}^T}{\tilde{X}^T}$, $\widetilde{F}^T \tilde{X} = 0,\ \tilde{X}^T \tilde{X} = I_d$. Now, it suffice to show, that the system S is also orthogonal to all gradients $\nabla \tilde{f}_j(\overline{x}), j = 1, \dots, r$, or equivalently

$$\widetilde{F}^T X = 0. \tag{5.1}$$

In order to show (5.1) let us consider the local diffeomorphism $\tilde{\Phi} \circ \Phi^{-1} : V \cap (0^r \times \mathbb{R}^d) \to V \cap (0^r \times R^d)$. Since for any $t(0^r, y) \in V, y \in \mathbb{R}^d,\ t \in \mathbb{R}$, $(0^r \in \mathbb{R}^{n-d})$ we have

$$\lim_{t \to 0} \frac{\tilde{\Phi} \circ \Phi^{-1}(t(0^r, y)) - \tilde{\Phi} \circ \Phi^{-1}(0)}{t} = \nabla(\tilde{\Phi} \circ \Phi^{-1})(0)(0^r, y)^T \in 0^r \times \mathbb{R}^d\ ,$$

it follows

$$\nabla(\tilde{\Phi} \circ \Phi^{-1})(0) \cdot \begin{pmatrix} 0 \\ I_d \end{pmatrix} = \nabla\tilde{\Phi}(\overline{x})(\nabla\Phi(\overline{x}))^{-1} \begin{pmatrix} 0 \\ I_d \end{pmatrix} = \begin{pmatrix} 0 \\ B \end{pmatrix} \tag{5.2}$$

with some (nonsingular) $d \times d$-matrix B. Putting $G := (F^T F)^{-1}$, we find using $\nabla\Phi(\overline{x}) = \binom{F^T}{X^T}$,

$$\begin{pmatrix} F^T \\ X^T \end{pmatrix} \cdot (FG\ X) = I_n \qquad \text{or} \qquad \nabla\Phi(\overline{x})^{-1} = (FG\ X).$$

Substituting into (5.2) gives

$$\begin{pmatrix} 0 \\ B \end{pmatrix} = \begin{pmatrix} \widetilde{F}^T \\ \tilde{X}^T \end{pmatrix} \cdot (FG\ X) \begin{pmatrix} 0 \\ I_d \end{pmatrix} = \begin{pmatrix} \widetilde{F}^T \\ \tilde{X}^T \end{pmatrix} X = \begin{pmatrix} \widetilde{F}^T X \\ \tilde{X}^T X \end{pmatrix},$$

i.e., (5.1).

□

In the equivalent definitions CM1, CM2, and CM3, a manifold M is locally defined around any $\overline{x} \in M$ by pairs $(U, \{f_j\}_{j=1}^r)$, (U, Φ), $(\tilde{U}, \tilde{\Phi})$ (called *charts*). The neigborhoods $U = U_{\overline{x}}, \tilde{U} = \tilde{U}_{\overline{x}}$ and functions $f_j, \Phi, \tilde{\Phi}$ depend on $\overline{x}$. Thus, the collection of all neighborhoods $U_{\overline{x}}$,

$$\bigcup_{\overline{x} \in M} U_{\overline{x}} \qquad \text{is an open cover of } M \text{ (possibly uncountable).} \tag{5.3}$$

When M is compact, then, we can choose a finite cover. But even in the case, that M is not compact we can choose a countable cover. It is well-known (see [32, p.1, 1.1.2 Def. and following observation]) that for a manifold M we have the equivalence:

M as a topological space, cf., Remark 5.1 has a countable base (is second countable)	$\Leftrightarrow$	M can be covered by countably many charts $(U_{x_\ell}, \Phi_{x_\ell})$.

Together with the observations in Remark 5.1 we obtain the

COROLLARY 5.1. *Let M be a manifold of $\mathbb{R}^n$. Then, in (5.3) we always can choose a countable cover, i.e., we can choose $x_\ell \in M,\ \ell \in \mathbb{N}$, and charts $(U_{x_\ell}, \Phi_{x_\ell})$ such that*

$$\bigcup_{\ell \in \mathbb{N}} U_{x_\ell} \quad \text{is an open cover of } M \ .$$

EX. 5.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ define a (global) C^k-diffeomorphism, $k \geq 1$. If M is a C^k-submanifold of $\mathbb{R}^n$ of codimension c_d, then $f(M)$ is also a C^k-manifold in $\mathbb{R}^n$ of the same codimension c_d.*
The same holds for a C^k-diffeomorphism $f : X \to Y$ between manifolds X, Y in $\mathbb{R}^n$.

Proof. Let near $\overline{x} \in M$ the manifold M be defined by the open neighborhood $U_{\overline{x}}$ and the C^k-functions $g_j : U_{\overline{x}} \to \mathbb{R}$, $j = 1, \ldots, c_d$, with linearly independent gradients $\nabla g_j(\overline{x})$. Since f is a diffeomorphism, $V := f(U_{\overline{x}})$ is an open neighborhood of $f(\overline{x})$. Indeed, f^{-1} is in particular continuous and thus the inverse image $V = (f^{-1})^{-1}(U_{\overline{x}}) = f(U_{\overline{x}})$ of $U_{\overline{x}}$ wrt. f^{-1} is open.
Moreover we have $y \in f(M) \cap V$ iff using $x = f^{-1}(y), \overline{y} = f(\overline{x})$, the equations

$$h_j(y) := g_j(f^{-1}(y)) = 0, \quad j = 1, \ldots, c_d \ ,$$

are satisfied, where the gradients $\nabla h_j(\overline{y}) = \nabla g_j(\overline{x}) \cdot \nabla f^{-1}(\overline{y})$ are linear independent due to the nonsingularity of $\nabla f^{-1}(\overline{y})$.

□

The following theorem is an immediate consequence of Definition 5.1.

THEOREM 5.2. *Let $S \subset \mathbb{R}^n$ be open and let $f : S \to \mathbb{R}^r$ be a C^k-function with $k \geq 1$. Suppose that for all $\overline{x} \in f^{-1}(0)$ the Jacobian $\nabla f(\overline{x})$ has full rank r $(r \leq n)$. Then, the inverse image $f^{-1}(0) = \{x \in S \mid f_j(x) = 0,\ j = 1, \ldots, r\}$ is a C^k-manifold of $\mathbb{R}^n$ of dimension $n - r$ (codimension r).*

Proof. Let $\overline{x} \in f^{-1}(0)$ and chose $U = S$ (independent of $\overline{x}$). Since $\nabla f(\overline{x})$ has full rank r, the r rows $\nabla f_j(\overline{x}), j = 1, \ldots, r$, must be linearly independent, i.e.. CM1 is satisfied.

□

Without the assumption in Theorem 5.2, that $\nabla f(\overline{x})$ must be of full rank, the set $f^{-1}(0)$ need not be a manifold. This fact becomes evident from the following Theorem due to Whitney (see, e.g., [33, IV.1.2 Lemma] for a proof).

THEOREM 5.3. *Let $S \subset \mathbb{R}^n$ be an arbitrary closed set. Then there exists a C^∞-function $f : \mathbb{R}^n \to \mathbb{R}$ such that $S = f^{-1}(0)$.*

The following implication of Corollary 5.1 is usefull.

THEOREM 5.4. *Any C^k-manifold $M \subset \mathbb{R}^n$ $(k \geq 1)$ of dimension $d < n$ has Lebesgue-measure zero in $\mathbb{R}^n$.*

Proof. By Corollary 5.1 the manifold M can be covered by countably many charts $(U_{x_\ell}, \Phi_{x_\ell})$, $\ell \in \mathbb{N}$. Thus, it suffice to prove the statement locally, i.e., it suffice to show that for any local diffeomorphism $\Phi_{x_\ell} : U_{x_\ell} \to V_\ell$ (ℓ fixed) as given in CM2 we have $\mu(M \cap U_{x_\ell}) = 0$. In view of CM2, $\Phi_{x_\ell}^{-1}$ is a C^1-function with $\Phi_{x_\ell}(U_{x_\ell} \cap M) = V_\ell \cap (0^{n-d} \times \mathbb{R}^d)$. Since $d < n$ by Example 2.1 the set $0^{n-d} \times \mathbb{R}^d$ has measure zero in $\mathbb{R}^n$, i.e., $\mu(\Phi_{x_\ell}(U_{x_\ell} \cap M)) = 0$. Then, by Theorem 2.1 it follows $\mu(\Phi_{x_\ell}^{-1}(\Phi_{x_\ell}(U_{x_\ell} \cap M))) = \mu(U_{x_\ell} \cap M) = 0$.

□

We finish the section with a more specific result. For a proof we refer to [21, Lemma 3.1.8].

THEOREM 5.5. *Let $M \subset \mathbb{R}^n$ be a C^k-manifold, $k \geq 1$. Then:*

1. *Every connected component $M_c \subset M$ is a C^k-manifold.*
2. *If M is connected then M is pathconnected.*

5.2. Stratifications and Whitney regular stratifications of sets in $\mathbb{R}^n$

In this section we consider special partitions of subsets in $\mathbb{R}^n$.

Stratification.

DEFINITION 5.3. *Let S be a subset of $\mathbb{R}^n$, and let $\Sigma = \{S_j,\ j \in J\}$ be a family of subsets $S_j \subset S,\ j \in J,\ J$ a finite index set.*

a) $\Sigma = \{S_j, j \in J\}$ *is called a partition of S if*

$$S = \bigcup_{j \in J} S_j \quad \textit{and} \quad S_j \cap S_\ell = \emptyset \ \textit{for all} \quad j, \ell \in J,\ j \neq \ell\ .$$

b) *A partition Σ of S is called a C^k-stratification ($k \geq 1$) of S if every set $S_j \in \Sigma, j \in J$, is a C^k-manifold.*
In this case, we call a pair (S, Σ) a stratified set in $\mathbb{R}^n$, and each $S_j \in \Sigma,\ j \in J$, a stratum of S.

As illustrative examples we consider four different partitions of $\mathbb{R}^2$.

1. $\Sigma_1 = \{S_1^0, S_2^2\}$ with $S_2^2 = \{0\}, S_1^0 = \mathbb{R}^2 \setminus \{0\}$
2. $\Sigma_2 = \{S_3^0, S_4^1\}$ with $S_4^1 = \{x \in \mathbb{R}^2 \mid x_1 \cdot x_2 = 0\}$, $S_3^0 = \mathbb{R}^2 \setminus S_4^1$
3. $\Sigma_3 = \{S_8^0, S_9^1, S_{10}^2\}$ with $S_{10}^2 = \{0\}$, $S_9^1 = \{x \in \mathbb{R}^2 \mid x_1 \cdot x_2 = 0\} \setminus \{0\}$, $S_8^0 = \mathbb{R}^2 \setminus (S_9^1 \cup S_{10}^2)$.
4. $\Sigma_4 = \{S_5^0, S_6^1, S_7^1\}$ with $S_7^1 = \{x \in \mathbb{R}^2 \mid x_1 = 0\} \setminus \{0\}$, $S_6^1 = \{x \in \mathbb{R}^2 \mid x_2 = 0\}$, $S_5^0 = \mathbb{R}^2 \setminus (S_7^1 \cup S_6^1)$.

The partitions 1,3,4 are stratifications of $\mathbb{R}^2$. Since S_4^1 is not a manifold (S_4^1 contains an intersection point at $x = 0$), the partition Σ_2 does not provide a stratification.

All stratifications considered lateron will arise from a concrete practical problem. The partitions above can be related to the following problems.

1. Asking for the rank of all real 2×1 matrices $X = \binom{x_1}{x_2} \in \mathbb{R}^2$, we have

$$\begin{array}{rcll} S_2^2 & = & \{X \in \mathbb{R}^2 \mid \text{rank } X = 0\} & \text{is a manifold of codimension 2 in } \mathbb{R}^2 \\ S_1^0 & = & \{X \in \mathbb{R}^2 \mid \text{rank } X = 1\} & \text{is (open) a manifold of codimension 0 in } \mathbb{R}^2. \end{array}$$

2. Consider all real 2×2 diagonal matrices $X = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} \in \mathbb{R}^2$. Then it follows

$$S_4^1 = \{X \in \mathbb{R}^2 \mid \det X = 0\}.$$

3. Let us investigate the rank of matrices $X = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} \in \mathbb{R}^2$. Then obviously,

$$\begin{array}{rcll} S_{10}^2 & = & \{X \in \mathbb{R}^2 \mid \text{rank } X = 0\} & \text{(codimension 2)}, \\ S_9^1 & = & \{X \in \mathbb{R}^2 \mid \text{rank } X = 1\} & \text{(codimension 1)}, \\ S_8^0 & = & \{X \in \mathbb{R}^2 \mid \text{rank } X = 2\} & \text{(codimension 0)}. \end{array}$$

To show that a partition $\Sigma = \{S_j,\ j \in J\}$ of a subet $S \subset \mathbb{R}^n$ is a stratification, the following principle is useful.

LEMMA 5.2. [Local transformation principle]
The partition $\Sigma = \{S_j,\ j \in J\}$ is a C^k-stratification $(k \geq 1)$ of S if the following holds:

> *For any $\overline{j} \in J$ and $\overline{x} \in S_{\overline{j}}$ there exist a C^k-diffeomorphism $\Phi = \Phi_{\overline{x}}$, $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ such that $\Phi(S_j) = S_j$, for all $j \in J$, and a $\mathbb{R}^n$-neighborhood V of $\Phi(\overline{x})$ such that $V \cap S_{\overline{j}}$ is a C^k-manifold.*

Proof. Under our assumptions let V be the neighborhood of $\overline{y} = \Phi(\overline{x})$ and let $g : V \to \mathbb{R}^r$ be the C^k-function defining locally near $\overline{y}$ the C^k-manifold $S_{\overline{j}}$ via: $y \in V \cap S_{\overline{j}}$ iff $g(y) = 0$. Then by $U = \Phi^{-1}(V)$, $f : U \to \mathbb{R}^r$, $f(x) := g(\Phi(x))$ $(y = \Phi(x))$, obviously we have: $x \in U \cap S_{\overline{j}}$ iff $f(x) = 0$. So, locally near $\overline{x}$ the function f defines the manifold $S_{\overline{j}}$.

□

This principle will allows us lateron to restrict ourselves to consider only elements $\overline{y} = \Phi(\overline{x})$ in especially simple (standard-) form.

Whitney Regular Stratification.

We now introduce so-called Whitney Regular Stratifications (see also [21, Section 7.5]). These are stratifications enjoying certain additional regularity conditions. We will make use of this concept only for C^∞-stratifications of algebraic sets. So, also in the present section we restrict ourselves to C^∞-manifolds.

Let $G_{n,d}$ $(d, n \in \mathbb{N}, d \le n)$ denote the set of all d-dimensional subspaces of $\mathbb{R}^n$. The sets $G_{n,d}$ are called Grassmann manifolds.
The set $G_{2,1}$ for example can be identified with the following arc of the unit circle (cf. Figure 5.1):

$$G_{2,1} \equiv \{x = (\cos\varphi, \sin\varphi) \mid \varphi \in [0, \pi)\} \ .$$

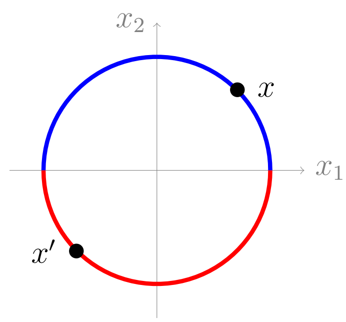

FIGURE 5.1. Grassmann manifold $G_{2,1}$.

By identifying antipodal points x, x' we can extend $G_{2,1}$ to be the whole unit circle. In this way $G_{2,1}$ becomes a compact manifold of dimension one. Such a simple construction is no more possible for $n \ge 3$. Note, that by identifying any d-dimensional subspace S of $\mathbb{R}^n$ with the $(n-d)$-dimensional orthogonal complement $S^\perp$, we can identify $G_{n,d}$ with $G_{n,n-d}$. In [21, p.314], by means of orthogonal projections the sets $G_{n,d}$ are presented as submanifolds of the sets $S^n(d)$ of symmetric $(n \times n)$-matrices of rank d.

THEOREM 5.6. *The set $G_{n,d}$ is diffeomorph to a compact submanifold of $S^n(d)$ $\quad (\subset \mathbb{R}^{(n+1)n/2})$ of dimension $(n-d)d$.*

DEFINITION 5.4. [Whitney regular stratification] *(See, e.g.,* [14, p.10], [21, Def. 7.5.1]*)*
Let (S, Σ) be a C^∞-stratified set in $\mathbb{R}^n$. Let X, Y be strata in Σ, $X \ne Y$.

(a) Let be given $\overline{x} \in X$. Then, Y is called Whitney regular over X at $\overline{x}$, if for every sequence of pairs $(x_j, y_j) \in X \times Y$, $j \in \mathbb{N}$, satisfying for $j \to \infty$,

(1) $x_j \to \overline{x}$, $y_j \to \overline{x}$
(2) $T_{y_j}Y \to T \subset G_{n,\dim Y}$
(3) $\mathcal{L}_j := span\{y_j - x_j\} \to \mathcal{L} \in G_{n,1}$,

the following holds: $\quad \mathcal{L} \subset T$.

(b) The stratum Y is said to be Whitney regular over X if Y is Whitney regular over X at every point $x \in X$. (S, Σ) is called a Whitney regular stratification if every stratum $Y \in \Sigma$ is Whitney regular over every stratum $X \in \Sigma$, $X \ne Y$.

As examples let us consider the stratifications 3, 4 above. In Σ_4 we choose $Y = S^1_7$, $X = S^1_6$, $\overline{x} = 0$, and the sequence $(x_j, y_j) = ((-\frac{1}{j}, 0), (0, \frac{1}{j}))$, $j = 1, 2, \ldots$ (see Figure 5.2). We find

$T_{y_j}Y = \{\lambda\binom{0}{1} \mid \lambda \in \mathbb{R}\} =: T$
$\mathcal{L}_j = \text{span}\{y_j - x_j\} = \{\lambda\binom{1}{1} \mid \lambda \in \mathbb{R}\} =: \mathcal{L}.$

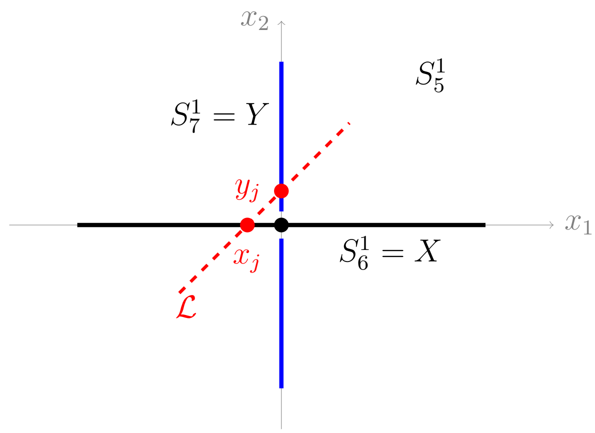

FIGURE 5.2. Statification with stratum $Y = S_7^1$ which is not Whitney regular over $X = S_6^1$ at $\overline{x} = 0$.

Obviously, $\mathcal{L} \not\subset T$ and Y is not Whitney regular over X at $\overline{x}$. Note that since $\overline{x} = 0 \notin Y$ the stratum X is Whitney regular over Y at 0. We invite the reader to prove the fact that the stratification Σ_3 of $\mathbb{R}^2$ (see Section 5.2.1) is Whitney regular.

EX. 5.3. *Let (S, Σ) be a Whitney regular stratification of the set $S \subset \mathbb{R}^m$. Then $(\mathbb{R}^n \times S, \mathbb{R}^n \times \Sigma)$ is a Whitney regular stratification of $\mathbb{R}^n \times S$ in $\mathbb{R}^{n+m}$ and $(\{0\} \times S, \{0\} \times \Sigma)$ is a Whitney regular stratification of $\{0\} \times S$.*

Proof. We only have to mention that for $\mathbb{R}^n$ (with only one stratum) the conditions in Definition 5.4 (Whitney regularity) trivially hold. Then for the product stratification $(\mathbb{R}^n \times S, \mathbb{R}^n \times \Sigma)$ the relations in Definition 5.4 are fulfilled by noticing that for any stratum Y in Σ we have $T_{(\overline{x},\overline{y})}(\mathbb{R}^n \times Y) \equiv T_{\overline{x}}(\mathbb{R}^n) \times T_{\overline{y}}(Y) = \mathbb{R}^n \times T_{\overline{y}}(Y)$. The same arguments are valid for $\{0\} \times S$.

□

REMARK 5.2.
(a) In view of condition (1) in Definition 5.4(a) it follows that Y is trivially Whitney regular over X at $\overline{x} \in X$ if $\overline{x} \notin \text{cl}\, Y$.

(b) The fact that Y is Whitney regular over X at $\overline{x} \in X$ remains invariant after some (local) diffeomorphism (see Ex. 5.2)

$$\Phi : U \to \mathbb{R}^n, \text{ with } \Phi(X \cap U) = X', \; \Phi(Y \cap U) = Y' \;,$$

where U is a neighbourhood of $\overline{x}$. This means that Y' is Whitney regular over X' at $x' = \Phi(\overline{x})$. (See [21, Rem. 7.5.2]).

Whitney regular stratifications enjoy certain nice properties. We will give these results without proofs.

THEOREM 5.7. *Let (S, Σ), $S \subset \mathbb{R}^n$, be a C^∞-stratified set. Suppose that $X, Y \in \Sigma$, $X \neq Y$, $\overline{x} \in X \cap \operatorname{cl} Y$ and Y is Whitney regular over X in $\overline{x}$. Then,*

(a) $\dim X < \dim Y$.

(b) If $\{y_j\}_{j=1}^{\infty} \subset Y,\ y_j \to \overline{x}, T_{y_j}Y \to T\ (\subset G_{n, \dim Y})$ then $T_{\overline{x}}X \subset T$.

Proof. Cf. [21, Lem. 7.5.7, Cor. 7.5.9]. □

THEOREM 5.8. *Let (S, Σ) be a Whitney regular stratification in $\mathbb{R}^n$.*

(a) Let S be closed and let Σ^j denote the union of all j-dimensional strata in Σ, $0 \leq j \leq n$. Then the sets

$$\bigcup_{j=0}^{k} \Sigma^j \qquad \text{are closed for all } k = 0, \ldots, n.$$

(b) Let S be locally closed. Let Σ^c denote the family of all connected components of the strata of Σ. Then,

1) *Σ^c is locally finite, i.e., for any $\overline{x} \in S$ there exists a $\mathbb{R}^n$-neighbourhood U such that $U \cap Y \neq \emptyset$ only for finitely many strata $Y \in \Sigma^c$.*
2) *[frontier conditions] If $X, Y \in \Sigma^c$ and $X \cap \operatorname{cl} Y \neq \emptyset$ then $X \subset \operatorname{cl} Y$ (i.e., $X \subset \partial Y$).*

Proof. See [21, Cor. 7.5.10] for (a), and [21, Th. 7.5.12], [14, p.16f] for (b). □

Whitney regular stratification of semi-algebraic sets.

THEOREM 5.9. *Each semi-algebraic set $S \subset \mathbb{R}^n$ allows (at least one) Whitney regular C^∞-stratification with finitely many semi-algebraic strata.*

Proof. See [14, p.12,(2.7)].

□

This general theorem allows another proof of the genericity result in Theorem 3.1.

EX. 5.4. *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a polynomial function, $p \not\equiv 0$. Then the inverse image $p^{-1}(0)$ is a (closed) set of measure zero.*

Proof. By Theorem 5.9 the algebraic set $p^{-1}(0)$ has a stratification into finitely many manifolds S_j, $j \in J$. It must follow,

$$\dim S_j \leq n - 1, \quad j \in J.$$

In fact, assuming $\dim S_{j_0} = n$ for some $j_0 \in J$, the stratum S_{j_0} is a (nonempty) open set in $\mathbb{R}^n$ and Ex. 2.5 implies $p \equiv 0$, a contradiction to $p \not\equiv 0$. Using Theorem 5.4 the result is proven.

□

Theorem 5.9 states that any semi-algebraic set allows a canonical Whitney-regular stratification. Note, that a given set can allow different Whitney-regular stratifications. Consider for example the stratifications Σ_1 and Σ_3 of $\mathbb{R}^2$ above. However, in the sequel we are interested in proving that a concrete partition of a given semi-algebraic subset $S \subset \mathbb{R}^n$ is a Whitney-regular stratification. The following homogeneity condition will be helpful.

DEFINITION 5.5. *Let be given a semi-algebraic set $S \subset \mathbb{R}^n$ and a semi-algebraic C^∞-stratification $\Sigma = \{S_j,\ j \in J\}$ of S, i.e., the C^∞-strata S_j are semi-algebraic (cf. Definition 2.4). The stratification Σ is said to satisfy the* homogeneity condition *if the following holds:*
Given any stratum $S_{j_0} \in \Sigma,\ j_0 \in J$, and $x, y \in S_{j_0}$ $(x \neq y)$, there exists R^n-neighbourhoods U of x and V of y and a C^∞-diffeomorphism $\Phi :\ U \to V,\ \Phi(x) = y$ which (locally) preserves the subsets $S_j \in \Sigma,\ j \in J$, i.e., $z \in S_j \cap U \Rightarrow \Phi(z) \in S_j \cap V$.

THEOREM 5.10. *Let be given a semi-algebraic set $S \subset \mathbb{R}^n$ and a semi-algebraic C^∞-stratification $\Sigma = \{S_j,\ j \in J\}$ of S. Suppose the stratification fulfills the homogeneity condition of Definition 5.5. Then, (S, Σ) is a Whitney-regular C^∞-stratified set.*

Proof. See [14]. □

5.3. Stratifications and Whitney regular stratifications of sets of matrices

In this section we investigate certain stratifications of sets of matrices. Let us recall some notations and introduce new ones. For given $n, m \in \mathbb{N}$, $m \geq n$, we define

$$\begin{aligned} M^{n,m} &= \{A \mid A \text{ is a real } (n \times m)\text{-matrix}\} \\ M^n &= M^{n,n} \\ S^n &= \{A \in M^n \mid A \text{ is symmetric}, \quad A^T = A\} \end{aligned}$$

It is obvious that the following identifications hold:

$$M^{n,m} \equiv \mathbb{R}^{n \times m}\ ,\quad S^n \equiv \mathbb{R}^N \text{ with } N = \frac{(n+1)n}{2}\ ,$$

Here , we consider stratifications of sets of matrices according to the ranks and define for fixed $n, m \in \mathbb{N}$, $m \geq n$,

$$\begin{aligned} M^{n,m}(j) &= \{A \in M^{n,m} \mid \text{rank } A = j\}, \quad j = 0, \dots, n, \\ S^n(j) &= \{A \in S^n \mid \text{rank } A = j\}, \quad j = 0, \dots, n, \end{aligned}$$

The following lemma will be used frequently.

LEMMA 5.3. *Let A be a real $(n \times m)$-matrix, decomposed as follows $(\ell > 0)$:*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \qquad \begin{matrix} A_{11} \text{ a } (\ell \times \ell)\text{-matrix}, & A_{12} \text{ a } (\ell \times k)\text{-matrix}, & n = \ell + p \\ A_{21} \text{ a } (p \times \ell)\text{-matrix}, & A_{22} \text{ a } (p \times k)\text{-matrix}, & m = \ell + k. \end{matrix}$$

Suppose, A_{11} is nonsingular. Then we have

$$\textit{rank } A = \ell \quad \Leftrightarrow \quad A_{22} = A_{21}\, A_{11}^{-1} A_{12}. \tag{5.4}$$

Proof. Since A_{11} is a nonsingular $(\ell \times \ell)$-matrix it follows that rank $\binom{A_{11}}{A_{21}} = \ell$. Consequently, rank $A = \ell$ iff there is a vector $x \in \mathbb{R}^\ell$ such that

$$\begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix} x = \begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix} \quad \text{or} \quad \begin{matrix} A_{11}x = A_{12} \\ A_{21}x = A_{22} \end{matrix} .$$

Then, A_{11} being regular, we find $x = A_{11}^{-1} A_{12}$ and $A_{22} = A_{21} A_{11}^{-1} A_{12}$.

□

Obviously, by

$$\Sigma = \{M^{n,m}(j), \quad j = 0, \ldots, n\} \tag{5.5}$$

we have given a partition of $M^{n,m}$. We will show that (5.5) yields a stratification.

THEOREM 5.11. *Let $n, m \in \mathbb{N}$ be given, $m \geq n \geq 1$. Then the sets $M^{n,m}(j)$ are C^∞-submanifolds of $M^{n,m}(\equiv \mathbb{R}^{n\times m})$ with*

$$\textit{codim } M^{n,m}(j) = (n-j)(m-j), \qquad j = 0, \ldots, n.$$

Thus, the partition Σ in (5.5) is a stratification of $M^{n,m}$.

Proof. Choose $j_0 \in \{0, \ldots, n\}$ and $\widehat{A} \in M^{n,m}(j_0)$. Then, $\widehat{A}$ contains a $(j_0 \times j_0)$-submatrix $\overline{A}_{11}$ of full rank j_0. We can choose appropriate permutation matrices P_1, P_2 such that $\overline{A} = P_1 \widehat{A} P_2$ has the form

$$\overline{A} = \begin{pmatrix} \overline{A}_{11} & \overline{A}_{12} \\ \overline{A}_{21} & \overline{A}_{22} \end{pmatrix}. \tag{5.6}$$

Since $\Phi : M^{n,m} \to M^{n,m}$, $\Phi(B) := P_1 B P_2$, is a strata preserving diffeomorphism, in view of the local transformation principle in Lemma 5.2 it suffice to show, that $M^{n,m}(j_0)$ fulfills the conditions of a manifold around $\overline{A} = \Phi(\widehat{A})$.
By continuity, there exists a $\mathbb{R}^{n\times m}$-neighborhood U of $\overline{A}$ such that for all $A \in U$ decomposed as in (5.6) the corresponding $(j_0 \times j_0)$-submatrix A_{11} is nonsingular. In view of Lemma 5.3, $A \in U$ has rank j_0 iff

$$f(A) := A_{22} - A_{21} A_{11}^{-1} A_{12} = 0. \tag{5.7}$$

To any element $a_{\rho\sigma}^{22}$ of the $(n - j_0) \times (m - j_0)$-matrix A_{22} there corresponds one rational (thus C^∞-) equation $f_{\rho\sigma}(A) = 0$, $\rho = 1, \ldots, n - j_0$, $\sigma = 1, \ldots, m - j_0$. Obviously for the partial derivatives we find

$$\frac{\partial}{\partial a_{\nu\mu}^{22}} f_{\rho\sigma}(A) = \begin{cases} 1 & \text{if } (\nu, \mu) = (\rho, \sigma) \\ 0 & \text{otherwise} \end{cases} .$$

Consequently, the derivatives $\nabla f_{\rho,\sigma}(\overline{A})$, $\rho = 1, \ldots, m - j_0$, $\sigma = 1, \ldots, m - j_0$, are linearly independent.

□

Ex. 5.5. *Let $\overline{A} \in M^n$ be a matrix in the manifold $M^n(n-1)$ of codimension 1. Then in a neigborhood $U_{\overline{A}}$ of $\overline{A}$ the following holds:*

$$A \in M^n \cap U_{\overline{A}} \text{ is in } M^n(n-1) \quad \Leftrightarrow \quad \det A = 0 \;.$$

Moreover, $\nabla_A \det(\overline{A}) \neq 0$.

Proof. We can assume that $\overline{A}$ has the form

$$\overline{A} = \begin{pmatrix} \overline{A}_{11} & \bar{b} \\ \bar{c}^T & \bar{\gamma} \end{pmatrix} \qquad \text{with } \overline{A}_{11}\text{, a nonsingular } (n-1)\times(n-1)\text{-matrix .}$$

By (5.7) the defining equation for the condition $A \in M^n(n-1) \cap U_{\overline{A}}$ for a matrix A of the form $A = \left(\begin{smallmatrix} A_{11} & b \\ c^T & \gamma \end{smallmatrix}\right)$ is given by $f(A) := \gamma - c^T A_{11}^{-1} b = 0$. In view of formula (7.2) the condition

$$\det A = \det A_{11} \cdot \det(\gamma - c^T A_{11}^{-1} b) = 0$$

is equivalent with (use $\det A_{11} \neq 0$) $\gamma - c^T A_{11}^{-1} b = f(A) = 0$. To show $\nabla_A \det(\overline{A}) \neq 0$ we only have to notice that by the multilinearity of $\det A$ we find

$$\frac{\partial}{\partial \gamma} \det \overline{A} = det \begin{pmatrix} \overline{A}_{11} & 0 \\ \bar{c}^T & 1 \end{pmatrix} = \det \overline{A}_{11} \neq 0 \;.$$

□

A result as in Theorem 5.11 is also valid for symmetric matrices.

Theorem 5.12. *Let $n \in \mathbb{N}$ be given. Then the sets $S^n(j)$ are submanifolds of S^n ($\equiv \mathbb{R}^N,\ N = (n+1)n/2$) with*

$$\textit{codim } S^n(j) = \frac{(n-j+1)(n-j)}{2}, \qquad j = 0, \ldots, n.$$

Thus the partition $\Sigma = \{S^n(j),\ j = 0, \ldots, n\}$ yields a stratification of S^n.

Proof. Let $j_0 \in \{0, \ldots, n\}$, $\overline{A} \in S^n(j_0)$, be given. Similar to the arguments in the proof of Theorem 5.11, by applying a transformation $P^T A P$, for any A in a neighborhood U of $\overline{A}$ we can assume a decomposition

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix}, \qquad \begin{array}{l} A_{22} \text{ a symmetric } (n-j_0)\times(n-j_0)\text{-matrix,} \\ A_{11} \text{ a nonsingular symmetric } (j_0 \times j_0)\text{-matrix.} \end{array}$$

Then, in view of Lemma 5.3 an element $A \in U \cap S^n$ is of rank j_0 iff

$$A_{22} - A_{12}^T A_{11}^{-1} A_{12} = 0.$$

Arguing as in the proof of Theorem 5.11, by symmetry of A_{22}, this identity represents $(n-j_0+1)(n-j_0)/2$ linearly independent equations.

□

As a corollary of Theorems 5.11,5.12, we obtain another proof of the result in Ex. 3.1.

Corollary 5.2. *Let be given $m, n \in \mathbb{N}$, $m \geq n \geq 1$. The set $M^{n,m}(n)$ $(S^n(n))$ of full rank n matrices is generic in $M^{n,m}$ (S^n).*

Proof. By Theorem 5.11 the set $M^{n,m}(n)$ is a manifold of codimension 0, thus open. Any of the manifolds $M^{n,m}(j)$, $0 \le j \le n-1$, has codimension ≥ 1, i.e., dim $M^{n,m}(j) <$ dim $M^{n,m} = nm$. By Theorem 5.4 these sets $M^{n,m}(j)$ are of measure zero. Consequently (cf. Lemma 2.1), $\bigcup_{j=0}^{n-1} M^{n,m}(j) = M^{n,m} \backslash M^{n,m}(n)$ has measure zero. The proof for $S^n(n)$ is similar.

$\square$

Theorems 5.11, 5.12 leads one to believe that a partition of any natural subspace S of $M^{m,n}$ (or S^n) according to the rank will always yield a stratification. Unfortunately, this is not the case in general. Problems arise if the elements of S have a special structure.
As an example we will examine symmetric band matrices. For given $n, q \in \mathbb{N}$, $1 \le q \le n$, we define the set $S^{n,q}$ of real symmetric $(n \times n)$-matrices of bandwidth q,

$$S^{n,q} = \{A \in S^n \mid a_{j\ell} = 0 \quad \text{for} \quad |j - \ell| \ge q\},$$

and

$$S^{n,q}(j) = \{A \in S^{n,q} \mid \text{ rank } A = j\}, \qquad j = 0, \ldots, n.$$

Obviously, the set $S^{n,q}$ can be identified with $\mathbb{R}^{N_q}$, where $N_q = n+(n-1)+\ldots+(n-q+1)$. We will make use of the following lemma.

LEMMA 5.4. *Let be given $A \in S^{n,q}$, $1 < q \le n$. Suppose, A has an eigenvalue ξ of multiplicity $m \ge q$. Then (at least) one of the elements in the q-th superdiagonal must be equal to zero, i.e.*

$$a_{1,q} \cdot a_{2,q+1} \ldots a_{n-q+1,n} = 0.$$

Proof. Consider the $(n - q + 1) \times (n - q + 1)$-submatrix

$$S = \begin{pmatrix} a_{1q} & 0 & \ldots & 0 \\ * & a_{2,q+1} & & \vdots \\ \vdots & & & 0 \\ * & \ldots\ldots & * & a_{n-q+1,n} \end{pmatrix}$$

of $(A - \xi I)$. Then it follows

$$\text{rank } S \le \text{rank}(A - \xi I) = n - m \le n - q.$$

This implies $\det S = a_{1,q} \cdot a_{2,q+1} \ldots a_{n-q+1,n} = 0$.

$\square$

Note, that by continuity the set $S^{n,q}(n)$ of nonsingular band matrices is open (and thus a manifold of codimension 0). However, for $1 \le j \le n-1$, the sets $S^{n,q}(j)$ need not be manifolds as will be shown by a simple counterexample.

Let us choose $n = 3$, $q = 2$ (tridiagonal matrices), $j = 1$, and

$$\overline{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in S^{3,2}(1).$$

Then we look for matrices A from $S^{3,2}(1)$,

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{12} & a_{22} & a_{23} \\ 0 & a_{23} & a_{33} \end{pmatrix} , \tag{5.8}$$

in a (small) neighborhood U of $\overline{A}$. Since A has 0 as double eigenvalue, by Lemma 5.3 we must have $a_{12}a_{23} = 0$. Hence, in view of (5.8), using $a_{22} \approx 1$, a matrix A is an element of $U \cap S^{3,2}(1)$ if and only if one of the following 3 cases occurs:

Case 0. $a_{12} = a_{23} = 0$ implying $a_{11} = a_{33} = 0$

Case 1. $a_{12} = 0$, $a_{23} \neq 0$ implying $a_{11} = 0$ and $a_{22}a_{33} - a_{23}^2 = 0$

Case 2. $a_{23} = 0$, $a_{12} \neq 0$ implying $a_{33} = 0$ and $a_{11}a_{22} - a_{12}^2 = 0$.

Thus,

$$A \in (M_0 \cup M_1 \cup M_2) \cap U \quad \Leftrightarrow \quad A \in S^{3,2}(1) \cap U$$

where

$$\begin{aligned} M_0 &= \{A \in S^{3,2}(1) \mid a_{12} = a_{23} = a_{11} = a_{33} = 0\} \\ M_1 &= \{A \in S^{3,2}(1) \mid f^1(A) = 0 \ , \ a_{23} \neq 0\} \quad \text{with} \quad f_1(A) = \begin{pmatrix} a_{12} \\ a_{11} \\ a_{22}\, a_{33} \ - a_{23}^2 \end{pmatrix} \\ M_2 &= \{A \in S^{3,2}(1) \mid f^2(A) = 0 \ , \ a_{12} \neq 0\} \quad \text{with} \quad f_2(A) = \begin{pmatrix} a_{23} \\ a_{33} \\ a_{11}\, a_{22} \ - a_{12}^2 \end{pmatrix} . \end{aligned}$$

An easy verification using $a_{22} \approx 1$ shows, that $\nabla f_1(\overline{A})$ and $\nabla f_2(\overline{A})$ have (full) rank 3. Consequently, M_1 and M_2 are C^∞-manifolds of dimension two in $S^{3,2} \equiv \mathbb{R}^5$, defined by different functions f_1, f_2 with different tangentspaces $T_{\overline{A}}M_1 \neq T_{\overline{A}}M_2$. The intersection

$$M_1 \cap M_2 = \{\alpha\overline{A} \mid \alpha \in \mathbb{R}\} = M_0$$

is an one-dimensional submanifold of M_1, M_2. This shows that $S^{3,2}(1)$ is not a manifold. However, this construction indicates, that by an appropriate sub-partitioning of the sets $S^{n,q}(j)$, a stratification (into manifolds) could be obtained.

We emphasize, that the stratification results in this section allow much deeper results than merely density results and statements on Lebesgue measures of certain strata as given in Corollary 5.2. Since the strata (submanifolds) are defined as solution sets of C^∞-functions, the tools of analysis and differential geometry can be applied. In Chapter 6 for example, we will apply Rene Thom's Transversality Theory to obtain so-called genericity results for parameter dependent families of matrices and corresponding problems.

Whitney regular stratifications of sets of matrices.

We now will prove the Whitney regularity of some stratifications of matrix sets.

THEOREM 5.13.

(a) The stratification $(M^{n,m}, \Sigma)$ $(m \geq n \geq 1)$ with $\Sigma = \{M^{n,m}(j),\ j = 0, \dots, n\}$ in Theorem 5.11 is Whitney regular.

(b) The stratification (S^n, Σ) with $\Sigma = \{S^n(j),\ j = 0, \dots, n\}$ in Theorem 5.12 is Whitney regular.

Proof. The proofs of (a) and (b) will be given simultaneously. Let $j \in \{0, \dots, n\}$ be fixed and put $X = M^{n,m} \equiv \mathbb{R}^{nm},\ X_j = M^{n,m}(j)$ for part (a) and $X = S^n \equiv \mathbb{R}^{(n+1)n/2}),\ X_j = S^n(j),\ m = n$ for part (b). We are going to show that the assumptions of Theorem 5.10 are satisfied.

X_j is a semi-algebraic subset of X: in view of $x^T A^T A x = \|Ax\|^2 \geq 0$ the matrix $A^T A$ is positive semidefinite and for such matrices the relation holds:

$$x^T A^T A x = 0 \quad \Leftrightarrow \quad Ax = 0 \quad \Leftrightarrow \quad A^T A x = 0 \ .$$

So the $(n \times m)$-matrix $A \in X$ has rank j, i.e., $A \in X_j$, iff the symmetric $(m \times m)$-matrix $A^T A$ has 0 as $(m - j)$-fold eigenvalue. In case (b) $A^T A$ can be replaced by the symmetric matrix A (with $m = n$) .

Now, consider the characteristic polynomial of $A^T A$,

$$\det(tI - A^T A) = t^m + c_m(A^T A) t^{m-1} + \cdots + c_1(A^T A).$$

Then, $A \in X_j$ iff $c_1(A^T A) = \cdots = c_{m-j}(A^T A) = 0$ and $c_{m-j+1}(A^T A) \neq 0$. From matrix theory it is known that the coefficients $c_k(A^T A)$ can be represented as the sum of all $(m + 1 - k) \times (m + 1 - k)$-principal minors of $A^T A$ (cf. e.g. [29, Section 4.11]). Consequently the mapping $\psi :\ X \to \mathbb{R}^{m-j+1},\ \psi_k(A) = c_k(A^T A),\ k = 1, \dots, m - j + 1,$ is a polynomial mapping such that

$$A \in X_j \ \Leftrightarrow\ \psi_k(A) = 0\ ,\ k = 1, \dots, m - j, \quad \psi_{m-j+1}(A) \neq 0.$$

The subset $V_j = \{x \in \mathbb{R}^{m-j+1} \mid x_1 = \cdots x_{m-j} = 0,\ x_{m-j+1} \neq 0\}$ of $\mathbb{R}^{m-j+1}$ is semi-algebraic. By construction, it follows $X_j = \psi^{-1}(V_j)$, which by Ex.2.8 is semi-algebraic.

The partition (X, Σ) fulfills the homogeneity propery (cf., Def.5.5): Let be given $A_1, A_2 \in X_j$ We have to construct a (local) strata preserving diffeomorphism $\Phi :\ U \to V$ from a X-neighbourhood U of A_1 onto a X-neighbourhood V of A_2 such that $\Phi(A_1) = A_2$. Since $A_1, A_2 \in X_j$, i.e., they have rank j, there exists the following transformation into standart-form:

In case (a) there exist nonsingular matrices C_1, C_2 in M^n, H_1, H_2 in M^m such that

$$C_k A_k H_k = \begin{pmatrix} I_j & 0 \\ 0 & 0 \end{pmatrix}, \qquad k = 1, 2\ ,$$

and in case (b) there exist nonsingular matrices B_1, B_2 in M^n such that

$$B_k A_k B_k^T = \begin{pmatrix} I_j & 0 \\ 0 & 0 \end{pmatrix}, \qquad k = 1, 2.$$

Then, the following mappings provide a (global) diffeomorphism $\Phi : X \to X$ which preserves the rank, i.e., the strata, and satisfies $\Phi(A_1) = A_2$.

In case (a): $\Phi :\ X = M^{n,m} \to X$ is defined by $\Phi(A) = C_2^{-1}C_1AH_1H_2^{-1}$.

In case (b): $\Phi :\ X = S^n \to X$ is defined by $\Phi(A) = B_2^{-1}B_1AB_1^T(B_2^T)^{-1}$.

□

In connection with linear equations $Ax = b$ with symmetric matrices A, we consider the set

(5.9) $$S1^n = \{B = (A\ b) \mid A \in S^n,\ b \in \mathbb{R}^n\} \quad \equiv \mathbb{R}^{N_1},\ \text{where } N_1 = (n+3)n/2\ .$$

Obviously, the following defines a partition of $S1^n$:

(5.10) $$\Sigma = \{S1^n_{j,0},\ j = 0, \dots, n\} \cup \{S1^n_{j,1},\ j = 0, \dots, n-1\}\ ,$$

where

$$\begin{aligned} S1^n_{j,0} &= \{B = (A\ b) \in S1^n \mid \text{rank } A = \text{rank } (A\ b) = j\}, \\ S1^n_{j,1} &= \{B = (A\ b) \in S1^n \mid \text{rank } A = j,\ \text{rank } (A\ b) = j+1\}. \end{aligned}$$

THEOREM 5.14. *The partition $(S1^n, \Sigma)$ of $S1^n$ defined by (5.10) is a Whitney regular stratification. The codimensions of the strata are given by:*

$$\text{codim } S1^n_{j,0} = \frac{(n-j)(n-j+3)}{2},\quad \text{codim } S1^n_{j,1} = \frac{(n-j)(n-j+1)}{2}.$$

Proof. We firstly apply Theorem 5.10 to show that the partition yields a Whitney regular stratification. The proof is similar to the proof of Theorem 5.13.

The sets $S1^n_{j,0}$, $S1^n_{j,1}$ are semi-algebraic: Let be given $B = (A\ b) \in S1^n$. Then

(5.11) $$\begin{aligned} B \in S1^n_{j,0} \quad &\Leftrightarrow \quad 0 \text{ is a } (n-j)\text{-fold eigenvalue of } A \\ &\qquad\quad \text{and a } (n-j+1)\text{-fold eigenvalue of } B^TB. \\ B \in S1^n_{j,1} \quad &\Leftrightarrow \quad 0 \text{ is a } (n-j)\text{-fold eigenvalue of } A \text{ and } B^TB. \end{aligned}$$

The coefficients $c_k,\ b_k$ of the characteristic polynomials of $A,\ B^TB$, resp.,

$$\det(tI - A) = t^n + c_n(A)t^{n-1} + \dots + c_1(A),$$

$$\det(tI - B^TB) = t^{n+1} + b_{n+1}(B^TB)t^n + \dots + b_1(B^TB)\ ,$$

resp., are sums of minors of $A,\ B^TB$ resp. Thus, the mapping (with $c_{n+1} = b_{n+2} = 1$) $\psi :\ S1^n \to \mathbb{R}^{2(n-j)+3}$, $\psi_k(B) = c_k(A),\ k = 1, \dots, n-j+1,\ \psi_{n-j+1+k}(B) = b_k(B^TB),\ k = 1, \dots, n-j+2,$ is a polynomial function of B. By (5.11) we have

$$\begin{aligned} B \in S1^n_{j,0} \quad &\Leftrightarrow \quad \psi_k(B) = 0,\ k = 1, \dots, n-j, n-j+2, \dots, 2(n-j)+2 \text{ and} \\ &\qquad\ \ \psi_k(B) \neq 0,\ k = n-j+1, 2(n-j)+3 \\ B \in S1^n_{j,1} \quad &\Leftrightarrow \quad \psi_k(B) = 0,\ k = 1, \dots, n-j, n-j+2, \dots, 2(n-j)+1 \text{ and} \\ &\qquad\ \ \psi_k(B) \neq 0,\ k = n-j+1, 2(n-j)+2 \end{aligned}$$

We define the sets

$$\begin{aligned} V_{j,0} &= \{x \in \mathbb{R}^{2(n-j)+3} \mid x_k = 0,\ k = 1, \ldots, n-j, n-j+2, \ldots, 2(n-j)+2 \\ &\qquad \text{and } x_k \neq 0,\ k = n-j+1, 2(n-j)+3\ \}, \\ V_{j,1} &= \{x \in \mathbb{R}^{2(n-j)+3} \mid x_k = 0,\ k = 1, \ldots, n-j, n-j+2, \ldots, 2(n-j)+1 \\ &\qquad \text{and } x_k \neq 0,\ k = n-j+1, 2(n-j)+2\ \}\ . \end{aligned}$$

These sets are semi-algebraic, and by construction $S1^n_{j,0} = \psi^{-1}(V_{j,0}),\ S1^n_{j,1} = \psi^{-1}(V_{j,1})$, which in view of Ex. 2.8 are semi-algebraic.

The partition $(S1^n, \Sigma)$ satisfies the homogeneity property: Let be given $B_1 = (A_1\ b_1) \in S1^n_{j,0}$. Since rank $A_1 = j$ there exists a matrix $C_1 \in M^n$ such that $C_1 A_1 C_1^T = \binom{I_j\ 0}{0^T 0}$. Then,

$$C_1(A_1\ b_1) \begin{pmatrix} C_1^T & 0 \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} I_j & 0 & c_1 \\ 0^T & 0 & d_1 \end{pmatrix}, \quad \text{with } \begin{pmatrix} c_1 \\ d_1 \end{pmatrix} = C_1 b_1\ , \tag{5.12}$$

where $c_1 \in \mathbb{R}^j, d_1 \in \mathbb{R}^{n-j}$. Since rank $A_1 =$ rank $(A_1\ b_1) = j$, it follows that $d_1 = 0$. Furthermore, we obtain

$$\begin{pmatrix} I_j & 0 & c_1 \\ 0^T & 0 & 0 \end{pmatrix} \begin{pmatrix} I_n & -c_1 \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} I_j & 0 & 0 \\ 0^T & 0 & 0 \end{pmatrix}. \tag{5.13}$$

The same transformation holds for another $B_2 = (A_2\ b_2) \in S1^n_{j,0}$ with corresponding C_2, c_2, and $d_2 = 0$. Consider the mapping

$$\begin{aligned} \Phi(A, b) &= C_2^{-1} C_1 (A\ b) \begin{pmatrix} C_1^T & 0 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} I_n & -c_1 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} I_n & -c_2 \\ 0^T & 1 \end{pmatrix}^{-1} \begin{pmatrix} C_2^T & 0 \\ 0^T & 1 \end{pmatrix}^{-1} \\ &= (\hat{A}\ \hat{b}). \end{aligned}$$

Then, since $\binom{I_n\ -c}{0^T\ 1}^{-1} = \binom{I_n\ c}{0^T\ 1}$, the matrix $\hat{A}$ is symmetric. Thus, we have given a diffeomorphism $\Phi:\ S1^n \to S1^n$. By construction, $\Phi(A_1\ b_1) = (A_2\ b_2)$ and Φ preserves the ranks, i.e. the strata.

Now, let be given $B_1 = (A_1\ b_1), B_2 = (A_2\ b_2) \in S1^n_{j,1}$. We again will perform a transformation of these matrices into a standard form. We consider B_1. Firstly, we can apply the transformation (5.12). Since now rank $(A_1\ b_1) =$ rank $A_1 + 1 = j + 1$ we must have $d_1 \neq 0$. The transformation as given in (5.13) yields

$$\begin{pmatrix} I_j & 0 & c_1 \\ 0^T & 0 & d_1 \end{pmatrix} \begin{pmatrix} I_n & -c_1 \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} I_j & 0 & 0 \\ 0^T & 0 & d_1 \end{pmatrix}\ .$$

Now, since $0 \neq d_1 \in \mathbb{R}^{n-j}$, we can assume $(d_1)_r \neq 0$, for an index $r,\ 1 \leq r \leq n-j$. Suppose $r \neq n-j$. By the permutation matrix P_1 obtained from I_n by interchanging the rows $j+r$ and n, we find

$$P_1 \begin{pmatrix} I_j & 0 & 0 \\ 0^T & 0 & d_1 \end{pmatrix} \begin{pmatrix} P_1^T & 0 \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} I_j & 0 & 0 \\ 0^T & 0 & \bar{d}_1 \end{pmatrix}\ . \tag{5.14}$$

such that $\alpha_1 := (\overline{d}_1)_{n-j} = (d_1)_r \neq 0$. By multiplication with $\left(\begin{smallmatrix} I_n & 0 \\ 0^T & \alpha_1^{-1} \end{smallmatrix}\right)$ from the right we obtain a matrix of the form of the right-hand matrix in (5.14) with $(\overline{d}_1)_{n-j} = 1$. Then, we successively subtract $(\overline{d}_1)_k$ -times the n^{th} row of (5.14) from the $(j+k)^{th}$ row and perform the corresponding column transformations for $k = 1, \ldots, n-j-1$. In matrix form these transformations read with an appropriate nonsingular $(n \times n)$-matrix H_1:

$$H_1 \begin{pmatrix} I_j & 0 & 0 \\ 0^T & 0 & \overline{d}_1 \end{pmatrix} \begin{pmatrix} H_1^T & 0 \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} I_j & 0 & 0 \\ 0^T & 0 & e_{n-j} \end{pmatrix}, \quad e_{n-j} \in \mathbb{R}^{n-j} . \tag{5.15}$$

The same transformation holds for B_2 with corresponding $C_2, P_2, H_2, d_2, \alpha_2$. Then, by

$$\Phi : \ S1^n \to S1^n, \quad \Phi(A\ b) = L_2^{-1} L_1 (A\ b) R_1 R_2^{-1},$$

where $L_k = H_k P_k C_k$ and

$$R_k = \begin{pmatrix} C_k^T & 0 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} I_n & -c_k \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} P_k^T & 0 \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0^T & \alpha_k \end{pmatrix}^{-1} \begin{pmatrix} H_k^T & 0 \\ 0^T & 1 \end{pmatrix}, \quad k = 1, 2,$$

we have defined a global diffeomorphism. Moreover, Φ is (rank-) strata-preserving with $\Phi(A_1\ b_1) = (A_2\ b_2)$.

<u>Proof of the formula for the codimensions:</u> We start with $S1^n_{j,0}$ and choose an element $\overline{B} = (\overline{A}\ \overline{b}) \in S1^n_{j,0}$. By using the above transformation Φ of $\overline{B}$ into the standard form (5.13) we can assume that $\overline{B}$ has already the form $\overline{B} = (\overline{A}\ \overline{b})$ with $\overline{A} = \left(\begin{smallmatrix} I_j & 0 \\ 0^T & 0 \end{smallmatrix}\right)$, $\overline{b} = 0$. Take any $B = (A\ b)$ in a small neighbourhood $\overline{U}$ of $\overline{B}$ composed as follows,

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix}, \quad A_{11} \text{ a } (j \times j)\text{-matrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} .$$

By continuity, A_{11} is nonsingular. By Lemma 5.3, rank $A = j \quad \Leftrightarrow \quad A_{22} = A_{12}^T A_{11}^{-1} A_{12}$. The condition rank $(A\ b) =$ rank A is equivalent with the existence of a solution $x = (x_1, x_2)$ of

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} .$$

Together we find that, $A_{12}x_2 = b_1 - A_{11}x_1$, $A_{12}^T x_1 + A_{12}^T A_{11}^{-1} A_{12} x_2 = b_2$, giving $A_{12}^T x_1 + A_{12}^T A_{11}^{-1}(b_1 - A_{11}x_1) = b_2$ or $b_2 = A_{12}^T A_{11}^{-1} b_1$. Thus, the condition rank $(A\ b) =$ rank $A = j$ is equivalent with the conditions

$$A_{22} = A_{12}^T A_{11}^{-1} A_{12} \quad \text{and} \quad b_2 = A_{12}^T A_{11}^{-1} b_1. \tag{5.16}$$

These relations represent (by symmetry of A) $N := (n-j+1)(n-j)/2 + (n-j) = (n-j+3)(n-j)/2$ equations $f_k(A, b) = 0, \ k = 1, \ldots, N$. As in the proof of Theorem 5.12 one can easily check that the gradients $\nabla f_k(\overline{A}, \overline{b}), \ k = 1, \ldots, N$, are linearly independent.

Now let be given $\overline{B} = (\overline{A}\ \overline{b}) \in S1^n_{j,1}$ in the standard form (5.15). Let us consider an element $B = (A\ b)$ in a small neighbourhood $\overline{U}$ of $\overline{B}$. Then since rank $(\overline{A}\ \overline{b}) = j+1$ we obviously have $j + 1 \leq$ rank $(A\ b) \leq$ rank $A + 1$ or $j \leq$ rank A. Consequently

$$(A\ b) \in S1^n_{j,1} \cap \overline{U} \quad \Leftrightarrow \quad \text{rank } A = j.$$

By Theorem 5.12 this defines (locally) a manifold of codimension $(n-j+1)(n-j)/2$.
$\square$

5.4. Transversal intersections of manifolds

If two manifolds M_1, M_2 of $\mathbb{R}^n$ intersect "nicely" we say that they are in general position or that they intersect transversally. Here is the formal definition.

DEFINITION 5.6. [Transversal intersection of M_1 and M_2]
Let M_1, M_2 be C^k-manifolds in $\mathbb{R}^n (k \geq 1)$ with codimensions r_1, r_2. We say that M_1 and M_2 intersect transversally at $\overline{x} \in M_1 \cap M_2$, if with the systems of C^k-functions $h_j : U \to \mathbb{R},\ j = 1, \ldots, r_1,\ g_\ell : U \to \mathbb{R},\ \ell = 1, \ldots, r_2$, respectively, defining in a neighbourhood U of $\overline{x}$ the manifolds M_1, M_2, respectively, the following holds:

(5.17) $\quad \nabla h_j(\overline{x}), \nabla g_\ell(\overline{x}),\ j = 1, \ldots, r_1,\ \ \ell = 1, \ldots, r_2$ *are linearly independent .*

We say that M_1 and M_2 intersect transversally (notation $M_1 \pitchfork M_2$) if M_1 and M_2 intersect transversally at every point $\overline{x} \in M_1 \cap M_2$. Instead of saying that M_1 and M_2 intersect transversally one often also says that M_1 and M_2 are in general position.

REMARK 5.3. Let M_1, M_2 be C^k-manifolds in $\mathbb{R}^n (k \geq 1)$ with codimensions r_1, r_2.
(a) If $M_1 \cap M_2 = \emptyset$ then by Definition 5.6 we have $M_1 \pitchfork M_2$. If $M_1 \cap M_2 \neq \emptyset$ and $M_1 \pitchfork M_2$ then, the intersection $M = M_1 \cap M_2$ defines submanifolds of M_1, M_2, and $\mathbb{R}^n$ of codimensions r_2, r_1, and $r_1 + r_2$, respectively.
(b) By Definition 5.6 the following holds: Suppose that codim $M_1 +$ codim $M_2 > n$, or equivalently dim $M_1 +$ dim $M_2 < n$. Then we have

$$M_1 \pitchfork M_2 \quad \Rightarrow \quad M_1 \cap M_2 = \emptyset.$$

(c) We note, that if (5.17) is satisfied, then, by continuity, the condition (5.17) holds for all $x \in (M_1 \cap M_2) \cap U$ with a (sufficiently) small $\mathbb{R}^n$-neighbourhood U of $\overline{x}$.

Transversal intersections can also be characterized in terms of conditions on tangent spaces.

LEMMA 5.5. *Let M_1, M_2 be manifolds in $\mathbb{R}^n$. Then the following conditions are equivalent:*
(a) $M_1 \pitchfork M_2$
(b) For all $\overline{x} \in M_1 \cap M_2$: $T_{\overline{x}} M_1 + T_{\overline{x}} M_2 = \mathbb{R}^n$
(c) For all $\overline{x} \in M_1 \cap M_2$: $N_{\overline{x}} M_1 \cap N_{\overline{x}} M_2 = \{0\}$

Proof. With the notations of Definition 5.6 it follows for $A := \{\nabla h_j(\overline{x}), \nabla g_\ell(\overline{x}),\ j = 1, \ldots, r_1, \ell = 1, \ldots, r_2\}$:

$$\begin{aligned} A^\perp &= \{\nabla h_j(\overline{x}), \nabla g_\ell(\overline{x}),\ j = 1, \ldots, r_1, \ell = 1, \ldots, r_2\}^\perp \\ &= \{\nabla h_j(\overline{x}), j = 1, \ldots, r_1\}^\perp \cap \{\nabla g_\ell(\overline{x}),\ \ell = 1, \ldots, r_2\}^\perp \\ &= T_{\overline{x}} M_1 \cap T_{\overline{x}} M_2. \end{aligned}$$

Consequently, the system A is a linearly independent set if and only if

$$n - (r_1 + r_2) = \text{dim } A^\perp = \text{dim } (T_{\overline{x}} M_1 \cap T_{\overline{x}} M_2)$$

or equivalently

$$\begin{aligned}\dim\,(T_{\overline{x}}M_1 + T_{\overline{x}}M_2) &= \dim\, T_{\overline{x}}M_1 + \dim\, T_{\overline{x}}M_2 - \dim\,(T_{\overline{x}}M_1 \cap T_{\overline{x}}M_2) \\ &= (n-r_1) + (n-r_2) - n + (r_1+r_2) = n\ .\end{aligned}$$

To show the equivalence of (b) and (c) note that (b) becomes

$$\mathbb{R}^n = (T_{\overline{x}}M_1 + T_{\overline{x}}M_2) = (N_{\overline{x}}M_1^{\perp} + N_{\overline{x}}M_2^{\perp}) = (N_{\overline{x}}M_1 \cap N_{\overline{x}}M_2)^{\perp}\ ,$$

or equivalently $(N_{\overline{x}}M_1 \cap N_{\overline{x}}M_2) = \{0\}$.

□

We give some illustrative examples.

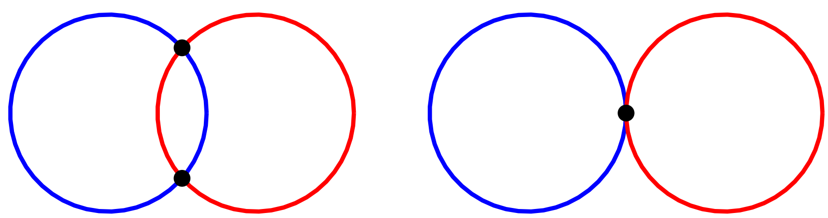

FIGURE 5.3. Two circles: transversal intersection (left), nontransversal intersection (right).

Figure 5.3 shows the intersection of two circles in $\mathbb{R}^2$. It is intuitively clear, that a nontransversal intersection of two manifolds M_1, M_2 can be transformed into a transversal one, by arbitrarily small perturbations. A nontransversal intersection can be regarded as an exceptional case. Moreover, a transversal intersection of two closed manifolds is stable wrt. small perturbations. However, if M_1 is closed and M_2 not, then, the "set of transversally intersecting M_1, M_2" need not be open. We give an example (cf. also [21, Remark 7.2.5]).

EX. 5.6. *Let be given the closed manifold $M_1 = \{0\}$, and the "relatively"open manifold $M_2 = \{(x_1, 0) \mid x_1 > 0\}$ in $\mathbb{R}^2$. Since $M_1 \cap M_2 = \emptyset$, by definition we have $M_1 \overline{\pitchfork} M_2$. However, under arbitrarily "small" perturbations of M_1 (or M_2) nontransversal intersections may appear, since* $\mathrm{cl}\, M_1 \cap \mathrm{cl}\, M_2 \neq \emptyset$ *(see Figure 5.4).*

FIGURE 5.4. Unperturbed manifolds M_1, M_2 intersecting transversally, perturbed manifolds M_1, $\tilde{M}_2$ show nontransversal intersection.

5.5. Transversality of mappings

In this section we again formulate all results for C^∞-manifolds and C^∞-functions, but we recall that the conclusions remain true for the C^k-case, $k \geq 1$.
Given a function $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ with components $(f_1, \ldots, f_m)$ we introduce the graph

$$\text{graph } (f) = \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R}^m \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^{n+m} .$$

Note that graph (f) is a C^∞-submanifold in $\mathbb{R}^{n+m}$ of codimension $c_d = m$, and dimension $d = n$ (see Definition 5.1). Indeed, $(x, y) \in$ graph (f) iff the system of m C^∞-equations,

$$g(x, y) := y - f(x) = 0$$

are satisfied. Since the Jacobian

$$\nabla g(x, y) = (-\nabla f(x) \; I_m)$$

has full rank m the gradients $\nabla g_j(x, y) = (-\nabla f_j(x), e_j)$, $j = 1, \ldots, m$, are linearly independent.

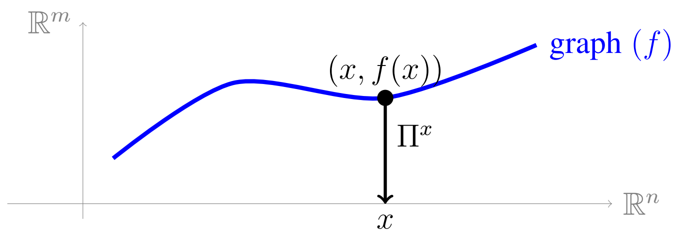

FIGURE 5.5. Sketch of the mapping graph (f).

Ex. 5.7. *Let $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$. The C^∞-mapping*

$$G_f : \mathbb{R}^n \to \text{graph } (f) \;\; , \qquad x \mapsto (x, f(x))$$

defines a C^∞-diffeomorphism between the manifolds of dimension n given by $\mathbb{R}^n$ and graph (f)*. Its inverse is given by*

$$G_f^{-1} = \Pi^x_{|\text{graph } (f)} \; ,$$

where $\Pi^x_{|\text{graph } (f)}$ is the restriction to graph (f) *of the projection $\Pi^x : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $\Pi^x(x, y) = x$.*

Proof. The mapping G_f is bijective and satisfies $(\Pi^x_{|\text{graph } (f)} \circ G_f)(x) = x$. □

Ex. 5.8.
(a) If $M \subset \mathbb{R}^m$ is a C^∞-manifold of codimension c_d, then $\mathbb{R}^n \times M$ is a C^∞-manifold in $\mathbb{R}^{n+m}$ of the same codimension c_d.
(b) Let M be a C^∞-submanifold of graph (f) *of codimension c_d, then $N = \Pi^x(M)$ is a C^∞-manifold in $\mathbb{R}^n$ of the same codimension c_d*

Proof. (a) Let the manifold M be locally defined in a neighborhood $V_{\overline{y}}$ of $\overline{y} \in M$ by the equations $g_j(y) = 0,\ j = 1, \ldots, c_d$. Then obviously for $(\overline{x}, \overline{y}) \in \mathbb{R}^n \times M$ an element $(x, y) \in \mathbb{R}^n \times V_{\overline{y}}$ is in $\mathbb{R}^n \times M$ iff the equations $h_j(x, y) := g_j(y) = 0, j = 1, \ldots, c_d$, are satisfied. By assumption the gradients $\nabla h_j(\overline{x}, \overline{y}) = (0, \nabla g_j(\overline{y}))$, $j = 1, \ldots, c_d$, are linearly independent.

(b) Recall that $G_f : \mathbb{R}^n \to \text{graph}\,(f)$ defines a C^∞-diffeomorphism. With $G_f^{-1}(M) = \Pi^x_{|\text{graph}\,(f)}(M) = N$, the statement follows by using Ex. 5.2.

□

DEFINITION 5.7. [f meets M transversally]

Let $M \subset \mathbb{R}^m$ be a manifold and let $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$. We say that f meets M transversally, notation $f \pitchfork M$, if the manifolds $M_1 = \text{graph}\,(f)$ and $M_2 = \mathbb{R}^n \times M$ intersect transversally, i.e., if $M_1 \pitchfork M_2$.

This definition implies the following.

LEMMA 5.6. *Let $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ and let M be a manifold in $\mathbb{R}^m$ of codimension c_d (dimension $d = m - c_d$). Assume $f \pitchfork M$ and $n < c_d$. Then*

$$\{f(x) \mid x \in \mathbb{R}^n\} \cap M = \emptyset\ ,$$

i.e., f does not meet M at all.

Proof. By Definition 5.7 the condition $f \pitchfork M$ means $\text{graph}\,(f) \pitchfork \mathbb{R}^n \times M$ for the manifold $M_1 := \text{graph}\,(f)$ of codimension m and $M_2 := \mathbb{R}^n \times M$ of codimension c_d in $\mathbb{R}^{n+m}$. The assumption $n < c_d$ implies

$$\text{codim}\ M_1 + \text{codim}\ M_2 = m + c_d > m + n\ ,$$

and Remark 5.3(b) yields $M_1 \cap M_2 = \emptyset$, or $\text{graph}\,(f) \cap (\mathbb{R}^n \times M) = \emptyset$, or equivalently

$$\{f(x) \mid x \in \mathbb{R}^n\} \cap M = \emptyset\ .$$

□

LEMMA 5.7. *Let $M \subset \mathbb{R}^m$ be a C^∞-manifold of codimension c_d, and $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$. Suppose $f \pitchfork M$. Then either $f^{-1}(M) = \emptyset$, or, $f^{-1}(M)$ is a manifold in $\mathbb{R}^n$ of codimension c_d.*

Proof. (See [21, Th.7.3.3]) Let $f \pitchfork M$ and $f^{-1}(M) \neq \emptyset$. Since $\text{graph}\,(f), \mathbb{R}^n \times M$ are manifolds in $\mathbb{R}^{n+m}$ of codimensions m, c_d, respectively, by Remark 5.3(a) the intersection $\text{graph}\,(f) \cap (\mathbb{R}^n \times M)$ defines a submanifold in $\text{graph}\,(f)$ of codimension c_d. In view of Ex. 5.2 (recall $\Pi^x_{|\text{graph}\,(f)}$ defines a diffeomorphism) the image $f^{-1}(M) = \Pi^x\big(\text{graph}\,(f) \cap (\mathbb{R}^n \times M)\big)$ is also a submanifold of $\mathbb{R}^n$ of codimension c_d.

□

We give a characterization of the transversal intersection of a mapping f with a manifold M in terms of the Jacobian of f and the tangent space of M.

THEOREM 5.15. *Let $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ and let $M \subset \mathbb{R}^m$ be a manifold. Then we have $f \pitchfork M$ if and only if for every $x \in \mathbb{R}^n$ with $f(x) \in M$ it holds,*

$$\nabla f(x)[\mathbb{R}^n] + T_{f(x)}M = \mathbb{R}^m \ .$$

Proof. (See [21, Th.7.3.4]) By definition and Lemma 5.5 the relation $f \pitchfork M$ holds iff for all $(\overline{x}, \overline{y})$ in graph $(f) \cap (\mathbb{R}^n \times M)$ we have:

$$T_{(\overline{x},\overline{y})}\text{graph } (f) + T_{(\overline{x},\overline{y})}(\mathbb{R}^n \times M) = \mathbb{R}^n \times \mathbb{R}^m \ . \tag{5.18}$$

Recall that (x, y) is in graph (f) iff $g(x, y) := y - f(x) = 0$. So, $T_{(\overline{x},\overline{y})}\text{graph } (f) = N_{(\overline{x},\overline{y})}(\text{graph } (f))^\perp = \ker(-\nabla f(\overline{x}) \ I_m)$. Obviously this set is spanned by the columns of

$$\begin{pmatrix} I_n \\ \nabla f(\overline{x}) \end{pmatrix} \ .$$

The tangent space $T_{(\overline{x},\overline{y})}(\mathbb{R}^n \times M)$ of the product $(\mathbb{R}^n \times M)$ is linearly isomorph to $T_{\overline{x}}\mathbb{R}^n \times T_{\overline{y}}M = \mathbb{R}^n \times T_{\overline{y}}M$. So (5.18) is valid iff the space $\nabla f(\overline{x})[\mathbb{R}^n]$ spanned by the columns of $\nabla f(\overline{x})$ satisfies $\nabla f(\overline{x})[\mathbb{R}^n] + T_{f(\overline{x})}M = \mathbb{R}^m$.

□

As an example consider the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{3}x^3 - x$. Take the manifold $M_c = \{c\}, c \in \mathbb{R}$, of codimension 1 in $\mathbb{R}$ and $T_cM_c = \{0\}$. Then by Theorem 5.15, for $c = 2/3$ the mapping f does not meet $M_{2/3}$ transversally, since at $\overline{x} = -1$ with $f(-1) \in M_{2/3}$ we have (cf., Figure 5.6)

$$f'(\overline{x})[\mathbb{R}] + T_cM_c = 0[\mathbb{R}] + \{0\} = \{0\} \neq \mathbb{R} \ .$$

For all $c \notin \{\pm 2/3\}$ we however find for any $\overline{x}$ with $f(\overline{x}) = c : \quad f'(\overline{x}) \neq 0$ und thus

$$f'(\overline{x})[\mathbb{R}] + T_cM_c = f'(\overline{x})[\mathbb{R}] = \mathbb{R} \ ,$$

and f meets M_c transversally.

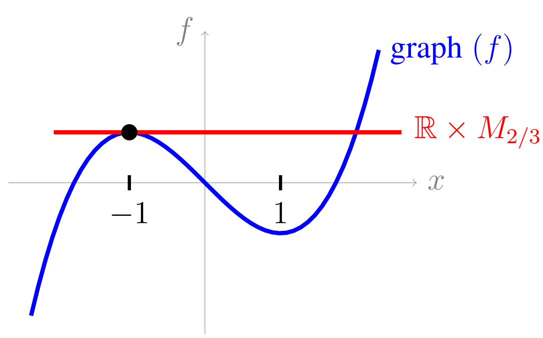

FIGURE 5.6. Nontransversal intersection of $M_{2/3}$ with f, or of $\mathbb{R} \times M_{2/3}$ with graph (f).

EX. 5.9. *Let $S \subset \mathbb{R}^m$ be a C^∞-manifold of codimension c_d and let $f \in C^\infty(\mathbb{R}^p, \mathbb{R}^m)$. Then $f \pitchfork S$ holds if and only if for every $\overline{x} \in \mathbb{R}^p$ with $f(\overline{x}) \in S$ the following is satisfied: if in a neighborhood $U_{\overline{y}}$ of $\overline{y} = f(\overline{x})$ the functions $h_j \in C^\infty(U_{\overline{y}}, \mathbb{R})$, $j = 1, \ldots, c_d$,* *define S (via $h_j(y) = 0$) then the functions $g_j(x) := h_j(f(x))$, $j = 1, \ldots, c_d$, are linearly independent in the sense,*

(5.19)
$$\nabla g_j(\overline{x}) = \nabla h_j(f(\overline{x})) \cdot \nabla f(\overline{x}),\ j = 1, \ldots, c_d, \quad \textit{are linearly independent vectors} .$$

Proof. Consider condition $f \pitchfork S$ or equivalently (cf., Definition 5.7) graph $(f) \pitchfork (\mathbb{R}^p \times S)$. The manifold graph (f) of codimension m in $\mathbb{R}^p \times \mathbb{R}^m$ is globally defined by the m equations $F(x, y) := y - f(x) = 0$ and near $\overline{y} = f(\overline{x})$ the manifold S of codimension c_d is given by the c_d equations

$$H(x, y) := \begin{pmatrix} h_1(y) \\ \vdots \\ h_{c_d}(y) \end{pmatrix} = 0 .$$

So, by definition, $f \pitchfork S$ means that the $m + c_d$ rows (gradients of the defining equations) of the $(m + c_d) \times (p + m)$-matrix

$$\begin{pmatrix} \nabla_x F(\overline{x}, \overline{y}) & \nabla_y F(\overline{x}, \overline{y}) \\ 0 & \nabla_y H(\overline{x}, \overline{y}) \end{pmatrix} = \begin{pmatrix} -\nabla f(\overline{x}) & I_m \\ 0 & \nabla h(\overline{y}) \end{pmatrix} \tag{5.20}$$

are linearly independent. This implies $c_d \leq p$. By applying Gaussian elimination, subtracting "$\nabla h(\overline{y})$ times the rows of $(-\nabla f(\overline{x})\ I_m)$ from the c_d last rows of $(0\ \nabla h(\overline{y}))$", the right-hand side of (5.20) becomes

$$\begin{pmatrix} -\nabla f(\overline{x}) & I_m \\ \nabla h(\overline{y}) \nabla f(\overline{x}) & 0 \end{pmatrix} ,$$

and this matrix has rank $c_d + m$ iff the matrix $\nabla h(\overline{y}) \nabla f(\overline{x})$ has full rank c_d.

□

CHAPTER 6

Genericity results for parametric problems

In this chapter we examine the genericity properties of parametric problems. Again, only C^∞-functions and C^∞-manifolds are considered. The results are based on so-called transversality theorems which are originally due to René Thom.
However, the presentation of the material and the proofs in Section 6.1,6.3,6.5 are essentially taken from the landmark book [21].

6.1. Jet transversality theorems

For later purposes we have to extend the notion of the graph for a function $f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$, $f(x) = (f_1(x), \ldots, f_m(x))$.
The 1-jet extension $j^1 f$ of f is given by the mapping

$$\begin{aligned} j^1 f : \mathbb{R}^n &\to \mathbb{R}^{N(n,m,1)} \quad \text{with} \quad N(n,m,1) = n + m + mn \\ x &\mapsto \Big[x, f_1(x), \ldots, f_m(x), \\ &\qquad \frac{\partial}{\partial x_1} f_1(x), \ldots, \frac{\partial}{\partial x_n} f_1(x), \frac{\partial}{\partial x_1} f_2(x), \ldots \ldots \frac{\partial}{\partial x_n} f_m(x) \Big] \end{aligned}$$

The 2-jet extension $j^2 f$ of f is the mapping

$$\begin{aligned} j^2 f : \mathbb{R}^n &\to \mathbb{R}^{N(n,m,2)} \quad \text{with} \quad N(n,m,2) = n + m + mn + m \cdot \frac{1}{2}(n+1)n \\ x &\mapsto \Big[j^1 f(x), \frac{\partial^2}{\partial x_1 \partial x_1} f_1(x), \frac{\partial^2}{\partial x_2 \partial x_1} f_1(x), \ldots, \frac{\partial^2}{\partial x_n \partial x_n} f_1(x), \\ &\qquad \frac{\partial^2}{\partial x_1 \partial x_1} f_2(x), \ldots, \frac{\partial^2}{\partial x_n \partial x_n} f_2(x), \ldots, \frac{\partial^2}{\partial x_1 \partial x_1} f_m(x), \ldots, \frac{\partial^2}{\partial x_n \partial x_n} f_m(x) \Big] \end{aligned}$$

In the same way we can define the ℓ-jet extension $j^\ell f$, $\ell \geq 3$, of f by

$$\begin{aligned} j^\ell f : \mathbb{R}^n &\to \mathbb{R}^{N(n,m,\ell)} \\ x &\mapsto \Big[j^{\ell-1} f(x) \;, \; \frac{\partial^\ell}{\partial x_1^\ell} f_1(x), \frac{\partial^\ell}{\partial x_2 \partial x_1^{\ell-1}} f_1(x), \ldots \;\; \ldots, \frac{\partial^\ell}{\partial x_n^\ell} f_m(x) \Big] \end{aligned}$$

where $N(n,m,\ell)$ is the length of the vector $j^\ell f(x)$. The space $J(n,m,\ell) := \mathbb{R}^{N(n,m,\ell)}$ is called the jet space of order ℓ. The function graph (f) coincides with the 0-jet extension $j^0 f$.
Note that $j^\ell f = \{ j^\ell f(x) \mid x \in \mathbb{R}^n \}$ is a manifold of dimension n in $J(n,m,\ell)$.

The following transversality theorems from Jongen et al. in [21] are the basis for proving genericity results for parametric problems.

THEOREM 6.1. [Jet Transversality Theorem]
Let be given $n, m \geq 1,\ \ell \geq 0$*, and a manifold* $M \subset J(n, m, \ell)$*. Then the set*

$$\mathcal{P} = \{f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \mid j^\ell f \pitchfork M\} \tag{6.1}$$

is generic (thus dense, cf., Definition 2.3) in $C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ *wrt. the* C^k_s*-topology for any* $k \geq \ell + 1$*. If moreover* M *is closed in* $J(n, m, \ell)$*, then the set* $\mathcal{P}$ *in* (6.1) *is also open in this topology for* $k \geq \ell + 1$*.*
If $f \in \mathcal{P}$ *and* $(j^\ell f)^{-1}(M) \neq \emptyset$ *then* $(j^\ell f)^{-1}(M)$ *is a manifold in* $\mathbb{R}^n$ *of codimension,* codim $(j^\ell f)^{-1}(M) =$ codim M.

Proof. We refer the reader to [21, Th.7.4.5] and [18, 2.1 Tranversality Theory] for a proof. The statement on the codimension of $(j^\ell f)^{-1}(M)$ follows from Lemma 5.7. □

This result can be extended to stratifications.

THEOREM 6.2. [Jet Transversality Theorem for stratifications]
Let be given $n, m \geq 1,\ \ell \geq 0$*, and a Whitney regular stratified set* (S, Σ)*,* $S \subset J(n, m, \ell)$*. Then the set*

$$\mathcal{P} = \{f \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \mid j^\ell f \pitchfork X \ \textit{ for all } \ X \in \Sigma\} \tag{6.2}$$

is generic (thus dense, cf., Definition 2.3) in $C^\infty(\mathbb{R}^n, \mathbb{R}^m)$ *wrt. the* C^k_s*-topology for any* $k \geq \ell + 1$*. If moreover the stratified set* S *is closed in* $J(n, m, \ell)$ *then the set* $\mathcal{P}$ *in* (6.2) *is also open in this topology for* $k \geq \ell + 1$*.*
Furthermore for $f \in \mathcal{P}$*, the inverse images* $\{(j^\ell f)^{-1}(X) \mid X \in \Sigma\}$ *form a Whitney regular stratification for* $(j^\ell f)^{-1}(S)$ *satisfying* codim $(j^\ell f)^{-1}(X) =$ codim X *(for all* X *such that* $(j^\ell f)^{-1}(X) \neq \emptyset$*).*

Proof. We refer the reader to [21, Th.7.5.11] and [18, 2.1 Tranversality Theory] for a proof. The statement on the codimension of $(j^\ell f)^{-1}(X)$ follows from Lemma 5.7. □

We often are not interested in the whole jet extension $j^\ell f$ but only in a part of it. This can be done with the help of

LEMMA 6.1. *Let* $F \in C^\infty(\mathbb{R}^n, \mathbb{R}^{m_1} \times \mathbb{R}^{m_2})$*. With a manifold* $\widetilde{M}$ *in* $\mathbb{R}^{m_2}$ *and an open set* S_0 *in* $\mathbb{R}^{m_1}$*, consider the manifold* $M := S_0 \times \widetilde{M}$ *in* $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$*. Let further with the projection* $\Pi_2(y_1, y_2) = y_2$*,* $(y_1, y_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$*, be defined* $\widetilde{F} = \Pi_2 \circ F$*. Then it holds,*

$$F \pitchfork S_0 \times \widetilde{M} \quad \Leftrightarrow \quad \widetilde{F} \pitchfork \widetilde{M}\ .$$

Proof. We make use of the formulas (Π_1 is the projection $\Pi_1(y_1, y_2) = y_1$) $\nabla \widetilde{F}(x) = \nabla(\Pi_2 \circ F)(x) = [\nabla F(x)]_{y_2}$ and

$$T_{F(x)}(S^0 \times \widetilde{M}) \equiv T_{\Pi_1(F(x))} S_0 \times T_{\widetilde{F}(x)} \widetilde{M} = \mathbb{R}^{m_1} \times T_{\widetilde{F}(x)} \widetilde{M}\ ,$$

to obtain that $F \pitchfork S_0 \times \widetilde{M}$ is equivalent with (see Theorem 5.15)

$$\nabla F(x)[\mathbb{R}^n] + T_{F(x)}(S_0 \times \widetilde{M}) = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \; ,$$

or

$$[\nabla F(x)]_{y_1}[\mathbb{R}^n] \times [\nabla F(x)]_{y_2}[\mathbb{R}^n] + \mathbb{R}^{m_1} \times T_{\widetilde{F}(x)}\widetilde{M} = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \; .$$

By taking the last m_2 relations we obtain our statement.

□

REMARK 6.1. As an example of Lemma 6.1 consider for $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$ the 1-jet extension

$$F(x) := j^1 f(x) = \big(x, f(x), \nabla f(x)\big) \; , \qquad J(n,1,1) = (\mathbb{R}^n \times \mathbb{R}) \times \mathbb{R}^n \; ,$$

with $m_1 = n+1, m_2 = n$ and the reduced 1-jet extension

$$\tilde{j}^1 f : \mathbb{R}^n \to \mathbb{R}^n, \;\; \tilde{j}^1 f(x) := \nabla f(x), \;\; \widetilde{M} = \{0\}, \;\; M = (\mathbb{R}^n \times \mathbb{R}) \times \widetilde{M} \; .$$

By Lemma 6.1 we have

$$j^1 f \pitchfork M \quad \Leftrightarrow \quad \tilde{j}^1 f \pitchfork \widetilde{M} \quad \Leftrightarrow \quad \nabla f \pitchfork \{0\} \; .$$

We wish to point out that the results below on parametric problems based on the Jet Transversality Theorems 6.1,6.2, in particular contain the genericity results for the corresponding non-parametric problems in Chapter 4 as special cases. See Remark 6.2 below for details.

6.2. Genericity results for parametric families of matrices

In this section we study parametric families $A(t)$ of matrices. In the first subsection we are interested in the generic behavior of such families wrt. the rank, and in the second subsection in the generic properties of $A(t)$ wrt. eigenvalues. In both cases we only consider families of symmetric matrices.

6.2.1. The ranks of parametric families of symmetric matrices.

In Corollary 5.2 we have seen that generically matrices $A \in S^n$ have full rank n. So generically, matrices $A \in S^n$ avoid all manifolds

$$S^n(j) \quad \text{of matrices of rank } j \le n-1 \; .$$

We now consider, families $A(t)$ of matrices in S^n depending smoothly on a parameter $t \in \mathbb{R}^p$. More precisely, let be given a function

$$A : \mathbb{R}^p \to S^n, \quad t \mapsto A(t) \; , \qquad A \in C^\infty(\mathbb{R}^p, S^n) \; . \tag{6.3}$$

We now cannot expect that generically $A(t)$ will be of rank n for all $t \in \mathbb{R}^p$.

Take, e.g., the family $A(t) \in S^2$, $t \in \mathbb{R}$, given by

$$A(t) = \begin{pmatrix} t & 1 \\ 1 & t \end{pmatrix} \; .$$

Then, the continuous function $\det A(t) = t^2 - 1$ has value -1 for $t_1 = 0$ and value 3 for $t_2 = 2$. Consequently by the Intermediate Value Theorem there must exist at least one point t_0 in the interval $(0, 2)$ such that $\det A(t_0) = 0$, i.e., $A(t_0)$ has rank < 2. In fact, $\det A(t_0) = 0$ for $t_0 = 1$ (or $t_0 = -1$) (see Figure 6.1). Moreover, any small smooth perturbation of $A(t)$ will result into a family of matrices which has rank < 2 for some t near $t_0 = 1$ (and near $t_0 = -1$).

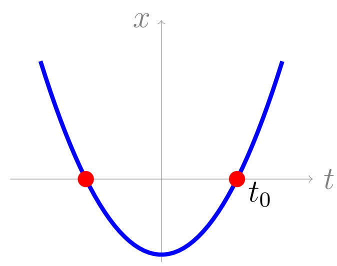

FIGURE 6.1. The function $\det A(t) = t^2 - 1$, with two zeroes $t_0 = \pm 1$.

We will now apply the Jet-Transversality Theorem 6.2 to examine which manifolds $S^n(j)$ are generically avoided by a matrix function $A(t)$. As can be expected the behavior essentially depends on the dimension p of the parameter space $\mathbb{R}^p$. Recall that by Lemma 5.6 for $A \pitchfork S^n(j)$, the family $A(t)$ does not meet $S^n(j)$ at all if $p <$ codim $S^n(j)$.

COROLLARY 6.1. [Genericity result for parametric families of symmetric matrices] *Let be given the Whitney regular stratified set (S^n, Σ) with $\Sigma = \{S^n(j)\ ,\ \ j = 0, \ldots, n\}$, see Theorem 5.13(b). Let*

$$\mathcal{P}_r = \{A \in C^\infty(\mathbb{R}^p, S^n) \mid A \,\overline{\pitchfork}\, S^n(j)\ \forall j = 0, \ldots, n\}\ .$$

Then the set $\mathcal{P}_r$ is dense and open in $C^\infty(\mathbb{R}^p, S^n)$ wrt. the C^k_s-topology for $k \geq 1$. Moreover, for $A \in \mathcal{P}_r$ the sets

$$S'_j = A^{-1}(S^n(j))\ , \qquad j = 0, \ldots, n\ ,$$

are manifolds in $\mathbb{R}^p$ which define a Whitney regular stratification $(\mathbb{R}^p, \Sigma')$ of $\mathbb{R}^p$ with strata set $\Sigma' = \{S'_j \mid j = 0, \ldots, n\}$.

Proof. To apply Theorem 6.2 to the matrix family in (6.3) we take the manifolds

$$M_j := \mathbb{R}^p \times S^n(j)\ , \qquad j = 0, \ldots, n\ .$$

By Ex.5.3 and Theorem 5.13(b) the family $\Sigma = \{M_j,\ j = 0, \ldots, n\}$ defines a Whitney regular stratification $(\mathbb{R}^p \times S^n, \Sigma)$ of $\mathbb{R}^p \times S^n$. Note that $S := \mathbb{R}^p \times S^n$ is a closed set. Then Theorem 6.2 states that for $\ell = 0$ and the jet-extension $j^0 A =$ graph (A) we have that the set

$$\mathcal{P}^*_r := \{A \in C^\infty(\mathbb{R}^p, S^n) \mid \text{graph } (A) \,\overline{\pitchfork}\, \mathbb{R}^n \times S^n(j)\ \forall j = 0, \ldots, n\}$$

is dense and open in the C_s^k-topology in $C^\infty(\mathbb{R}^p, S^n)$ for all $k \geq 1$. Recalling that by Definition 5.7 we have

$$A \pitchfork S^n(j) \quad \Leftrightarrow \quad \text{graph } (A) \pitchfork (\mathbb{R}^p \times S^n(j))$$

this proves the first statement.
Furthermore, Theorem 6.2 asserts, that for $A \in \mathcal{P}_r^*$ the sets $S_j^* := (\text{graph } (A))^{-1}(\mathbb{R}^p \times S^n(j))$ form a Whitney regular stratification of the set $(\text{graph } (A))^{-1}(\mathbb{R}^p \times S^n) = \mathbb{R}^p$. In view of

$$\begin{aligned}(\text{graph } (A))^{-1}(\mathbb{R}^p \times S^n(j)) &= \{t \in \mathbb{R}^p \mid (t, A(t)) \in \mathbb{R}^p \times S^n(j)\} \\ &= \{t \in \mathbb{R}^p \mid A(t) \in S^n(j)\} = A^{-1}(S^n(j))\end{aligned}$$

also the second claim follows.

□

To illustrate this genericity result let

$p = 1$, i.e., $t \in \mathbb{R}$: Suppose A is in $\mathcal{P}_r$. By Lemma 5.6 we have

$$\{A(t) \mid t \in \mathbb{R}\} \cap S^n(j) = \emptyset \quad \text{if} \quad 1 < \text{codim } S^n(j) \; .$$

For $j = n-1$ we have codim $S^n(n-1) = 1$ and for $j \leq n-2$ it follows codim $S^n(j) \geq$ codim $S^n(n-2) = 3$ (cf., Theorem 5.12). Consequently in our case $p = 1$, the set $\{A(t) \mid t \in \mathbb{R}^p\}$ does generically not meet any set $S^n(j)$ of matrices of rank $j \leq n-2$.
For the manifold $S^n(n-1)$ we possibly have $A^{-1}(S^n(n-1)) \neq \emptyset$ but by Lemma 5.7 (for $A \in \mathcal{P}_r$) the set $A^{-1}(S^n(n-1))$ is a subset of $\mathbb{R}$ of codimension 1, i.e., this set consists of isolated points $t_i \in \mathbb{R}$, $i \in I$, with a possibly infinite index set I.
In other words, generically, a one parametric family $A(t),\ t \in \mathbb{R}$, of matrices in S^n only meets the manifold $S^n(n-1)$ on a discrete point set.

For the case $p > 1$, $t \in \mathbb{R}^p$: Here we can conclude in the same way, that $A \in \mathcal{P}_r$ does not meet $S^n(j)$ if:

$$p < \text{codim } S^n(j) = \frac{1}{2}(n-j+1)(n-j) \; .$$

Similar results are valid for families $A(t)$ of matrices in M^n (see, [36] for details).

6.2.2. Eigenvalues of parametric families of symmetric matrices.

In Section 4.2 we have shown that generically a real symmetric matrix $A \in S^n$ only has simple eigenvalues. In this section we are interested in the generic behavior of the eigenvalues of a parameter family of real symmetric matrices,

$$A : \mathbb{R}^p \to S^n \; , \quad t \mapsto A(t), \quad A \in C^\infty(\mathbb{R}^p, S^n) \; .$$

Here n and p are given natural numbers.

The genericity results below are based on the following stratification of S^n wrt. the distribution of multiplicities of eigenvalues. Let $A \in S^n$ have l distinct eigenvalues $\lambda_1 < \ldots < \lambda_l$ and let $m_j(A)$ denote the multiplicity of eigenvalue $\lambda_j,\ j = 1, \ldots, l$. We define

$$\sigma(A) = \{m_1(A), \ldots, m_l(A)\} \; .$$

Then $\sigma(A)$ characterizes the multiplicities of the eigenvalues of A. Now we introduce partitions σ of n into positive integers, $m_j,\ j = 1, \ldots, k$,

$$\sigma = \{m_1, \ldots, m_k\}\ , \qquad \sum_{j=1}^{k} m_j = n\ .$$

The set of all such partitions σ is denoted by $\mathcal{S}$. For $\sigma \in \mathcal{S}$ we can define the subsets

$$A_\sigma = \{A \in S^n \mid \sigma(A) = \sigma\} \subset S^n \equiv \mathbb{R}^{(n+1)n/2}\ .$$

Here is the stratification result.

THEOREM 6.3. [Stratification of S^n according to multiplicities of eigenvalues]
The partition $\Sigma = \{A_\sigma,\ \sigma \in \mathcal{S}\}$ forms a Whitney regular C^∞-stratification of the set S^n. Moreover, for $\sigma = \{m_j,\ j = 1, \ldots, k\} \in \mathcal{S}$ we have,

$$\text{codim } A_\sigma = \sum_{j=1}^{k} \left(\frac{(m_j + 1)m_j}{2} - 1 \right)\ .$$

Proof. For a proof of this result we refer to [25]. The proof uses techniques as in the proof of Theorem 5.13.

□

For a similar stratification result for real skew-symmetric matrices we refer to [26].

As a direct consequence of Theorem 6.2, for $\ell = 0$ and the 0-jet mapping $j^0 A = \text{graph } (A)$, we obtain the following genericity result.

COROLLARY 6.2. *Let (S^n, Σ) be the Whitney regular stratification of Theorem 6.3. Then the set*

$$\mathcal{P}_r = \{A \in C^\infty(\mathbb{R}^p, S^n) \mid A \,\overline{\pitchfork}\, A_\sigma \text{ for all } A_\sigma \in \Sigma\ \}$$

is dense and open in $C^\infty(\mathbb{R}^p, S^n)$ wrt. the C_s^∞-topology for all $k \geq 1$. Moreover for $A \in \mathcal{P}_r$ the sets $A^{-1}(A_\sigma), \sigma \in \mathcal{S}$, form a Whitney regular stratification (S^n, Σ'), $\Sigma' := \{A^{-1}(A_\sigma),\ \sigma \in \mathcal{S}\}$, of the set $\mathbb{R}^n = A^{-1}(S^n)$, with codim $A^{-1}(A_\sigma) =$ codim A_σ.

As a general observation for $A \in \mathcal{P}_r$ we note:

- Generically (i.e., for $A \in \mathcal{P}_r$) the family $A(t),\ t \in \mathbb{R}^p$, does not meet any set A_σ with codim $A_\sigma > p$.

The case $p = 1$ is especially interesting.

One-parametric families of symmetric matrices.

In this case $p = 1$, or $t \in \mathbb{R}$, generically all sets A_σ with multiple eigenvalue matrices are avoided by $A(t)$. Indeed, according to Theorem 6.3, the stratum A_{σ_0} with $\sigma_0 = \{1, \ldots, 1\}$ containing all matrices $A \in S^n$ with n distinct eigenvalues has codimension 0, i.e., the set A_{σ_0} is an open subset of S^n. The next stratum $A_{\sigma_1},\ \sigma_1 = \{2, 1, \ldots, 1\}$ with exactly one double and $n - 2$ simple eigenvalues has codimension

$$\text{codim } A_{\sigma_1} = \frac{3 \cdot 2}{2} - 1 = 2\ .$$

This leads to the following special case of Corollary 6.2.

COROLLARY 6.3. *Let $p = 1$ and let $A \in \mathcal{P}_r$. Then for all parameter values $t \in \mathbb{R}$ the matrices $A(t)$ have n simple eigenvalues.*
More precisely, the set of eigenvalues $\lambda_1(t), \ldots, \lambda_n(t)$ of $A(t)$ is given by n C^∞-functions $\lambda_j(t)$, $j = 1, \ldots, n$, $t \in \mathbb{R}$, which never intersect on the whole set $\mathbb{R}$ (see Figure 6.2).

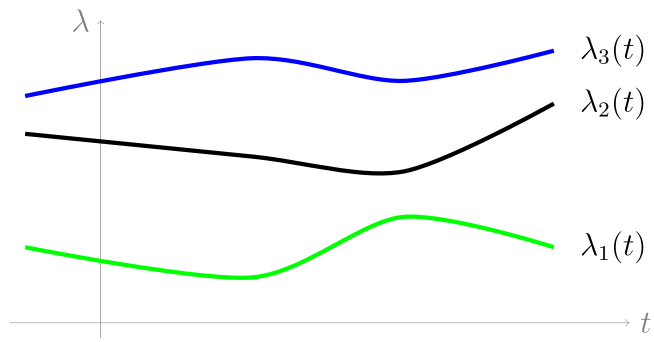

FIGURE 6.2. Eigenvalue curves of a generic one-parametric family $A(t)$.

The density part of Corollary 6.3 means, that if we have given a (non-generic) matrix function $A \in C^\infty(\mathbb{R}, S^n)$ with intersecting eigenvalue functions, then by abritrarily small appropriate C^∞-perturbations $E(t)$ (small wrt. the C^∞_s-topology) we obtain a family $\tilde{A}(t) = A(t) + E(t)$ from $\mathcal{P}_r$ with n non-intersecting eigenvalue curves $\tilde{\lambda}_j(t),\ t \in \mathbb{R}$.

Here a question arises: Can such a perturbation $E(t)$ be given explicitly?

As we shall see, the answer is yes, at least for one-parameter families $A(t)$ which are not only C^∞-functions but analytic functions of $t \in \mathbb{R}$.
Recall that a matrix function $A : \mathbb{R} \to S^n$, $A \in C^\infty(\mathbb{R}, S^n)$, is called analytic on $\mathbb{R}$, if at each $\bar{t} \in \mathbb{R}$, for any component $a_{ij}(t)$ of $A(t)$ the Taylor expansion

$$a_{ij}(t) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!} \frac{d^\nu}{dt^\nu} a_{ij}(\bar{t})(t - \bar{t})^\nu$$

is convergent in some interval $(\bar{t} - \bar{\varepsilon}, \bar{t} + \bar{\varepsilon})$ for some $\bar{\varepsilon} > 0$, $\bar{\varepsilon} = \varepsilon_{ij}(\bar{t})$, $i, j = 1, \ldots, n$. To understand why we need analyticity of $A(t)$, we have to summarize some facts on the stability of eigenvalues and eigenvectors of a one-parametric family of symmetric matrices from the landmark book of Kato [28].

Let $A \in C^\infty(\mathbb{R}, S^n)$. It is easy to show, by applying the Implicit Function Theorem, that near a simple eigenvalue $\overline{\lambda}$ of $A(\bar{t})$ there are locally defined C^∞-functions $\lambda(t),\ \lambda(\bar{t}) = \overline{\lambda}$, and $x(t)$, such that for t near $\bar{t}$ the value $\lambda(t)$ is a simple eigenvalue of $A(t)$ with corresponding eigenvector $x(t)$.
Problems arrize if $\overline{\lambda}$ is an eigenvalue with multiplicity $m > 1$. Then the eigenvalue functions are still C^∞-functions, but the corresponding eigenvectors need even not behave

continuously. The classical counterexample is due to Rellich (cf., e.g., [28, p. 111]),

$$A(t) = e^{1/t^2} \begin{pmatrix} \cos 2/t & \sin 2/t \\ \sin 2/t & -\cos 2/t \end{pmatrix}, \quad t \neq 0, \quad A(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This C^∞-matrix family (not analytic) has C^∞-eigenvalue functions $\lambda_{1,2}(t) = \pm e^{-1/t^2}$ with double eigenvalue $\lambda_{1,2}(0) = 0$ for $\bar{t} = 0$. But it can be shown that no corresponding eigenvector functions $x_{1,2}(t)$ can be defined as continuous functions near $\bar{t}$.
As with magic, things change when instead of merely a C^∞-function, the family $A(t)$ is analytic. Then we have the following deep result (cf., [28, Vol. II, Theorem 6.1]).

THEOREM 6.4. *Let $A : \mathbb{R} \to S^n$ be analytic on $\mathbb{R}$. Then, appropriately defined eigenvalue functions $\lambda_j(t)$ and corresponding orthonomal eigenvectors $x_j(t)$, $j = 1, \ldots, n$ (i.e., $\|x_j(t)\| = 1,\ x_j(t)^T x_i(t) = 0,\ j \neq i$), are analytic.*

Now, in case the eigenvector functions $x_j(t)$ are smooth (say C^∞- or analytic functions), then they allow a simple perturbation of $A(t)$ which completely avoids multiple eigenvalues.

We present an illustrative example,

$$A(t) = \begin{pmatrix} t & 0 \\ 0 & -t \end{pmatrix}, \quad t \in \mathbb{R},$$

with eigenvalue functions $\lambda_1(t) = t$, $\lambda_2(t) = -t$, intersecting at $\bar{t} = 0$ and corresponding (constant) eigenvectors $x_1(t) = e_1,\ x_2(t) = e_2$.
Perturbing the family A with the symmetric matrix $e_1 e_2^T + e_2 e_1^T$ (dyadic product with the eigenvectors e_1, e_2) leads to a function

$$\tilde{A}(t) = A(t) + \varepsilon\big(e_1 e_2^T + e_2 e_1^T\big) = \begin{pmatrix} t & \varepsilon \\ \varepsilon & -t \end{pmatrix}, \quad \varepsilon \neq 0,$$

with no more intersecting eigenvalue functions $\lambda^\varepsilon_{1,2}(t) = \pm\sqrt{t^2 + \varepsilon^2}$ (cf., Figure 6.3).

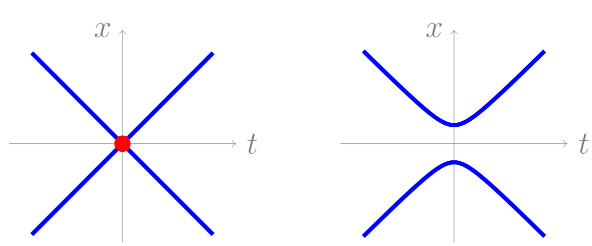

FIGURE 6.3. Eigenvalue functions of the unperturbed and the perturbed matrix family.

This perturbation can be generalized (see [36] for details).

THEOREM 6.5. *Let the matrix function $A : \mathbb{R} \to S^n$ be analytic on $\mathbb{R}$ and let $\lambda_j(t), x_j(t)$, $j = 1, \ldots, n$, be the analytic eigenpair functions according to Theorem 6.4, with possibly intersecting eigenvalue functions $\lambda_j(t)$. Then for any $\varepsilon \neq 0$ the perturbed analytic family*

$$A^\varepsilon(t) = A(t) + \varepsilon \sum_{j=1}^{n-1} \big(x_j(t) x_{j+1}(t)^T + x_{j+1}(t) x_j(t)^T \big)$$

has eigenvalue functions $\lambda_j^\varepsilon(t)$ which never intersect on the whole $\mathbb{R}$.

Proof. (See also [36, Corollary 1].) With the analytic orthogonal matrices $Q(t) = [x_1(t) \ldots x_n(t)]$ the families $A(t), A^\varepsilon(t)$ are transformed as follows:

$$Q(t)^T A(t) Q(t) = \begin{pmatrix} \lambda_1(t) & & 0 \\ & \ddots & \\ 0 & & \lambda_n(t) \end{pmatrix} ,$$

and

$$Q(t)^T A^\varepsilon(t) Q(t) = \begin{pmatrix} \lambda_1(t) & \varepsilon & & \cdots & 0 \\ \varepsilon & \lambda_2(t) & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \varepsilon \\ 0 & \cdots & \cdots & \varepsilon & \lambda_n(t) \end{pmatrix} =: T^\varepsilon(t) ,$$

where $\varepsilon \neq 0$. Obviously for any $\lambda \in \mathbb{R}$ the matrix $\big(T^\varepsilon(t) - \lambda I\big)$ has rank $\geq n-1$, since it contains a nonsingular $(n-1) \times (n-1)$-submatrix. So, the eigenvalues $\lambda_j^\varepsilon(t)$ of $T^\varepsilon(t)$ and thus of $A^\varepsilon(t)$ are simple for all $t \in \mathbb{R}$.

□

As a final example we take the matrix family

$$A(t) = \begin{pmatrix} t^2 & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & -t \end{pmatrix} , \quad t \in \mathbb{R} ,$$

with eigenvalue functions $\lambda_1(t) = t^2$, $\lambda_2(t) = t$, $\lambda_3(t) = -t$, and two double eigenvalues $\overline{\lambda} = 1$ at $\overline{t} = \pm 1$ and an eigenvalue $\overline{\lambda} = 0$ at $\overline{t} = 0$ with multiplicity 3 (see Figure 6.4). The eigenvalue curves of the perturbed family ($\varepsilon \neq 0$)

$$A^\varepsilon(t) = A(t) + \varepsilon\big(e_1 e_2^T + e_2 e_1^T + e_2 e_3^T + e_3 e_2^T\big) = \begin{pmatrix} t^2 & \varepsilon & 0 \\ \varepsilon & t & \varepsilon \\ 0 & \varepsilon & -t \end{pmatrix} , \quad t \in \mathbb{R} ,$$

never intersect (see Figure 6.5).

Let us remark that we also obtain similar results for the real eigenvalues of matrix families $A(t)$ in M^n. For the one-parametric case we refer to [36].

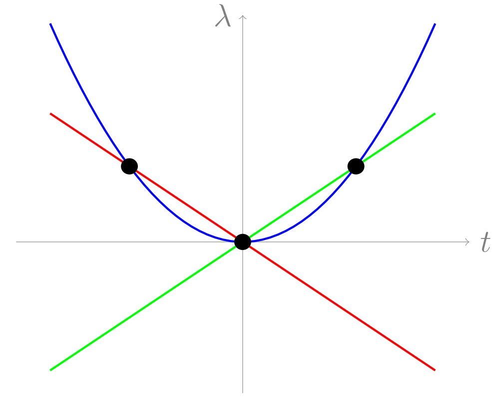

FIGURE 6.4. Three eigenvalue curves with two double eigenvalues and one triple eigenvalue of the family $A(t)$.

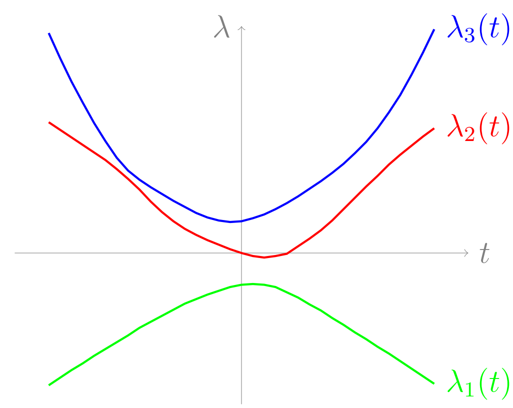

FIGURE 6.5. Eigenvalue curves of the perturbed matrix family $A^\varepsilon(t)$ without intersections.

6.3. Genericity results for parametric unconstrained programs

Given a function $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R})$, we consider problems

$$P(t): \qquad \min_{x\in\mathbb{R}^n} f(x,t) \,, \tag{6.4}$$

depending on a parameter $t \in \mathbb{R}^p$. For any fixed $\bar{t}$ the program $P(\bar{t})$ represents an unconstrained minimization problem (see Section 4.3). In what follows, the numbers n, p are always fixed. The first subsection studies the general case $p \geq 1$, whereas the second subsection especially looks at one-parametric problems. We again emphasize that the material presented in this section is essentially taken from [21, Section 10.1,10.2].

6.3.1. Parametric unconstrained programs.

Recall from Section 4.3 that a local minimizer $\overline{x}$ of $P(\bar{t})$ must necessarily be a solution of the critical point equation

$$\nabla_x^T f(\overline{x}, \bar{t}) = 0 \,. \tag{6.5}$$

If at such a point $(\overline{x}, \overline{t})$ the regularity condition,

$$\nabla_x^2 f(\overline{x}, \overline{t}) \quad \text{is nonsingular,} \tag{6.6}$$

is fulfilled, then by the Implicit Function Theorem 7.2, around $(\overline{x}, \overline{t})$, the solution set of (6.5) is a C^∞-manifold in $\mathbb{R}^n \times \mathbb{R}^p$ of dimension p parameterized by a (locally defined) C^∞-function $x(t)$ with $x(\overline{t}) = \overline{x}$. We again call a point $(\overline{x}, \overline{t})$ satisfying (6.5) and (6.6) a nondegenerate critical point.
For a non-parametric unconstrained minimization problem given by $f(x)$ we have proven in Section 4.3 that generically all critical points of f are nondegenerate. Again here in the parametric case, we cannot expect that generically for all $t \in \mathbb{R}^p$ at all critical points (x, t), i.e., at all points satisfying $\nabla_x^T f(x, t) = 0$, the relation (6.6) holds. However, as we shall see in Lemma 6.2, the following weaker condition will be satisfied generically:

$$\nabla_{(x,t)} \nabla_x^T f(x, t) \text{ has full rank } n \text{ at all solutions } (x, t) \text{ of } \nabla_x^T f(x, t) = 0\ . \tag{6.7}$$

Also under this condition the critical point set

$$C(f) = \{(x, t) \in \mathbb{R}^n \times \mathbb{R}^p \mid \nabla_x^T f(x, t) = 0\} \tag{6.8}$$

is a C^∞-manifold of dimension p (codimension n). Indeed, the set $C(f)$ is defined globally by the n equations $\nabla_x^T f(x, t) = 0$, which under (6.7) satisfy the conditions CM1 for a manifold (see Section 5.1). We call a point $(\overline{x}, \overline{t}) \in C(f)$ satisfying rank $\nabla_{(x,t)} \nabla_x^T f(\overline{x}, \overline{t}) = n$ but rank $\nabla_x^2 f(\overline{x}, \overline{t}) < n$ a turning point of the set $C(f)$.

LEMMA 6.2. *The set*

$$\mathcal{P}_r^1 = \{f \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}) \mid f \text{ fulfills } (6.7)\} \tag{6.9}$$

is open and dense (thus generic) in $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R})$ wrt. the C_s^k-topology, $k \geq 2$. Moreover, for $f \in \mathcal{P}_r^1$, the set $C(f)$ is a manifold of $\mathbb{R}^n \times \mathbb{R}^p$ of codimension n.

Proof. Consider the reduced j^1-function $\tilde{j}^1 f(x, t) = \nabla_x f(x, t)$. By Theorem 5.15 the condition $\tilde{j}^1 f \pitchfork \{0\}$ holds iff for any solution (x, t) of $\nabla_x^T f(x, t) = 0$ we have (use $T_{\nabla_x f(x,t)}\{0\} = \{0\}$, $\tilde{j}^1 f = \nabla_x f$)

$$\nabla_{(x,t)} \nabla_x^T f(x, t)[\mathbb{R}^n \times \mathbb{R}^p] + \{0\} = \mathbb{R}^n\ ,$$

or equivalently $\nabla_{(x,t)} \nabla_x^T f(x, t)$ has full rank n. Consequently, $\tilde{j}^1 f \pitchfork \{0\}$ iff $f \in \mathcal{P}_r^1$. According to Lemma 6.1 (see also Remark 6.1) $\tilde{j}^1 f \pitchfork \{0\}$ holds iff $j^1 f \pitchfork (\mathbb{R}^n \times \mathbb{R}^p \times \{0\})$. Since the set $(\mathbb{R}^n \times \mathbb{R}^p \times \{0\})$ is closed, the statement now follows from the Jet Transversality Theorem 6.1.
The fact that for $f \in P_r^1$ the set $C(f)$ is a manifold has been shown before the lemma. □

This lemma in particular states that generically (for $f \in \mathcal{P}_r^1$) the critical point set $C(f)$ consists of a C^∞-manifold in $\mathbb{R}^n \times \mathbb{R}^p$ of codimension n. But now we are interested in the set of degenerate critical points, i.e., the points $(x, t) \in C(f)$ with singular Hessian $\nabla_x^2 f(x, t)$:

$$C_s(f) = \{(x, t) \in \mathbb{R}^n \times \mathbb{R}^p \mid \nabla_x^T f(x, t) = 0,\ \det \nabla_x^2 f(x, t) = 0\}\ . \tag{6.10}$$

To prove the main genericity result of this section, we consider the reduced 2-jet function

$$\tilde{j}^2 f(x,t) = \big(\nabla_x f(x,t), \nabla_x^2 f(x,t)\big) \in \mathbb{R}^n \times S^n \ ,$$

and the manifolds

$$M_j = \{0\} \times S^n(j) \ .$$

Note that the set $C_s(f)$ consists of all points (x,t) where $\tilde{j}^2 f$ meets one of the manifolds $M_j,\ j = 0,\dots,n-1$, i.e., points where $\nabla_x^T f(x,t) = 0$ but $\nabla_x^2 f(x,t)$ has rank $j \le n-1$. In view of Theorem 5.13(b) and Ex. 5.3 the stratification

$$(\{0\} \times S^n, \Sigma) \ , \qquad \text{with} \qquad \Sigma = \{M_j,\ j = 0,\dots,n\} \ ,$$

is Whitney regular. Also notice that the set $\{0\} \times S^n$ is closed. So, by applying Theorem 6.2 and Remark 6.1, in our context, we obtain the following genericity theorem for the set

$$\mathcal{P}_r = \{f \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}) \mid \tilde{j}^2 f \pitchfork M_j \ \forall M_j \in \Sigma\} \ . \tag{6.11}$$

THEOREM 6.6. *The set $\mathcal{P}_r$ is open and dense (thus generic) in $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R})$ wrt. the C^k_s-topology, $k \ge 3$. Moreover, for $f \in \mathcal{P}_r$, the inverse images*

$$(\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(j)\big) \ , \quad j = 0,\dots,n \ ,$$

are submanifolds in $C(f)$ of codimension $c_d = \text{codim } S^n(j) = \frac{1}{2}(n-j+1)(n-j)$, and

$$\Sigma^{-1} := \{(\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(j)\big) \mid j = 0,\dots,n\} \ ,$$

yields a Whitney regular stratification of the set $C(f) = (\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n\big)$.

Note that we have

$$f \in \mathcal{P}_r \qquad \Rightarrow \qquad f \in \mathcal{P}_r^1 \ . \tag{6.12}$$

Indeed, using Lemma 6.1 (with $S_0 \times \widetilde{M}$ replaced by $\widetilde{M} \times S_0$) with the open subset $S_0 = S^n(n)$ of S^n and $\widetilde{M} = \{0\}$ we have

$$\tilde{j}^2 f \pitchfork M_n = \{0\} \times S^n(n) \quad \Rightarrow \quad \nabla_x f \pitchfork \{0\} \ ,$$

where according to the proof of Lemma 6.2 $\nabla_x f \pitchfork \{0\}$ is equivalent with $f \in \mathcal{P}_r^1$.

We present some examples: Let $f \in \mathcal{P}_r$. Then the set $C(f) \subset \mathbb{R}^n \times \mathbb{R}^p$ is a manifold of codimension n, dimension p in $\mathbb{R}^n \times \mathbb{R}^p$. Further:

- $(\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(n)\big)$ is a relatively open submanifold of codimension n in $\mathbb{R}^n \times \mathbb{R}^p$ and of codimension 0 in $C(f)$.
- $(\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(n-1)\big)$ is a submanifold of codimension $n+1$ in $\mathbb{R}^n \times \mathbb{R}^p$ and of codimension 1 in $C(f)$.
- $(\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(n-2)\big)$ is a submanifold of codimension $n+3$ in $\mathbb{R}^n \times \mathbb{R}^p$ and of codimension 3 in $C(f)$.

  In particular, for $p \le 2$ the set $(\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(n-2)\big)$ is empty (for $f \in \mathcal{P}_r$).

So, roughly speaking, for "low" p the mapping $\tilde{j}^2 f$ only meets sets in $\{0\} \times S^n(j)$ with "hight" rank j (low codimension). More precisely for $f \in \mathcal{P}_r$ we have

$$p < \text{codim } S^n(j) \;\Rightarrow\; \{(x,t) \in \mathbb{R}^n \times \mathbb{R}^p \mid \nabla_x^T f(x,t) = 0,\; \nabla_x^2 f(x,t) \in S^n(j)\} = \emptyset \;.$$

We finish the section with some simple examples for $n = p = 1$. Consider

$$P(t): \qquad \min_{x \in \mathbb{R}} f(x,t) := \frac{1}{3}x^3 - tx \;.$$

The critical point set

$$C(f) = \{(x,t) \in \mathbb{R} \times \mathbb{R} \mid \nabla_x f(x,t) = x^2 - t = 0\}$$

is a $p = 1$ dimensional manifold in $\mathbb{R}^2$ (see Figure 6.6). Here, $f \in \mathcal{P}_r^1$ holds since

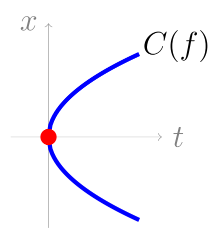

FIGURE 6.6. Critical point set $C(f)$ for $f(x,t) = \frac{1}{3}x^3 - tx$ with turning point $(\overline{x}, \overline{t}) = (0,0)$.

$$\nabla_{(x,t)} \nabla_x^T f(x,t) = (2x, -1) \qquad \text{has full rank } 1 \;.$$

This function is even in $\mathcal{P}_r$ with only one point $(\overline{x}, \overline{t})$ in the set

$$\begin{aligned} (\tilde{j}^2 f)^{-1}\big(\{0\} \times S^n(n-1)\big) &= \{(x,t) \mid \nabla_x f(x,t) = x^2 - t = 0,\; \nabla_x^2 f(x,t) = 2x = 0\} \\ &= \{(0,0)\} \;. \end{aligned}$$

So, $(\overline{x}, \overline{t}) = (0,0)$ is a so-called turning point of $C(f)$.

For the function $f(x,t) = \frac{1}{3}x^3 - t^2 x$ the citical point set reads

$$C(f) = \{(x,t) \mid x^2 - t^2 = 0\}$$

and $C(f)$ is not a manifold (see Figure 6.7).

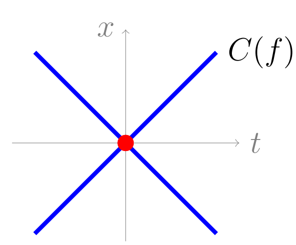

FIGURE 6.7. Critical point set $C(f)$ for $f(x,t) := \frac{1}{3}x^3 - t^2 x$, $f \notin \mathcal{P}_r$.

So, $f \notin \mathcal{P}_r^1$, and at the critical point $(\overline{x}, \overline{t}) = (0,0)$ we obtain

$$\nabla_{(x,t)} \nabla_x^T f(\overline{x}, \overline{t}) = (0,0) \ .$$

However, if we take the perturbations of f given by

$$f^{\pm}(x,t) := f(x,t) \pm \varepsilon x \ , \qquad \varepsilon > 0 \ ,$$

we obtain functions in $\mathcal{P}_r^1$ (even in $\mathcal{P}_r$). The set $C(f^-)$ (see Figure 6.8) only consists of nondegenerate critical points, i.e., points $(\overline{x}, \overline{t})$ such that $\det \nabla_x^2 f(\overline{x}, \overline{t}) \neq 0$ holds. The set $C(f^+)$ (see Figure 6.8) has two degenerate critical points $(\overline{x}, \overline{t}) = (0, \pm\sqrt{\varepsilon})$ which are turning points.

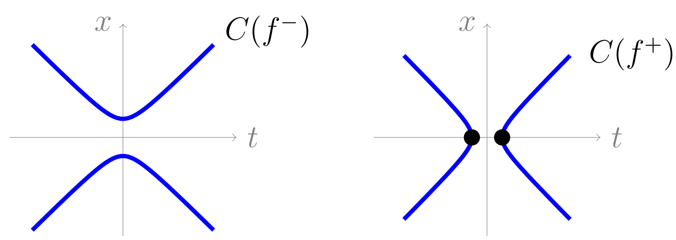

FIGURE 6.8. Critical points of the perturbed functions $f(x,t) \pm \varepsilon x$.

6.3.2. One-parametric unconstrained programs.

As an important special case of (6.4) we examine one-parametric families of problems $P(t)$, i.e., the case $p = 1$, $t \in \mathbb{R}$. We refer to [21, Section 10.2] for further details.

Ex. 6.1. *Let $(\overline{x}, \overline{t}) \in \mathbb{R}^n \times \mathbb{R}$ be a point in $C(f)$ such that $\nabla \nabla_x^T f(\overline{x}, \overline{t})$ ($\nabla = \nabla_{(x,t)}$) has rank n and $\det \nabla_x^2 f(\overline{x}, \overline{t}) = 0$, i.e., $(\overline{x}, \overline{t}) \in C_s(f)$. Then it holds:* rank $\nabla_x^2 f(\overline{x}, \overline{t}) = n-1$. *In particular, the latter condition is fulfilled for all $(\overline{x}, \overline{t}) \in C_s(f)$ if f is in $\mathcal{P}_r$.*

Proof. In view of $\nabla \nabla_x^T f(\overline{x}, \overline{t}) = \left(\nabla_x^2 f(\overline{x}, \overline{t}) \mid \frac{\partial}{\partial t} \nabla_x^T f(\overline{x}, \overline{t}) \right)$ and $\det \nabla_x^2 f(\overline{x}, \overline{t}) = 0$ we must have

$$\text{rank } \nabla_x^2 f(\overline{x}, \overline{t}) \leq n-1 \quad \text{and} \quad n = \text{rank } \nabla \nabla_x^T f(\overline{x}, \overline{t}) \leq \text{rank } \nabla_x^2 f(\overline{x}, \overline{t}) + 1 \ .$$

This gives rank $\nabla_x^2 f(\overline{x}, \overline{t}) = n-1$. Note that by (6.12), $f \in \mathcal{P}_r$ implies $f \in \mathcal{P}_r^1$ (cf., (6.9)), and thus for all $(\overline{x}, \overline{t}) \in C_s(f)$ the assumptions in Ex.6.1 are fulfilled. □

As we will see in a moment, there is a strong connection between the set $C_s(f)$ (cf., (6.10)) and the following optimization problem with objective $\phi(x,t) := t$,

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}} t \qquad \text{s.t.} \qquad \nabla_x^T f(x,t) = 0 \ . \tag{6.13}$$

We introduce the Lagrangean function L for this program,

$$L(x,t,\lambda) = \phi(x,t) + \sum_{i=1}^{n} \lambda_i \frac{\partial}{\partial x_i} f(x,t) \ ,$$

and recall that by Definition 6.1(b) below, a feasible point $(\overline{x},\overline{t})$ is a critical point of (6.13), if $\nabla\nabla_x^T f(\overline{x},\overline{t})$ has full rank n (LICQ holds), and with some multiplier vector $\overline{\lambda}\in\mathbb{R}^n$ we have ($\nabla=\nabla_{(x,t)}$),

$$\nabla^T L(\overline{x},\overline{t},\overline{\lambda}) = \begin{pmatrix}0\\1\end{pmatrix} + \sum_{i=1}^n \overline{\lambda}_i \nabla^T \frac{\partial}{\partial x_i} f(\overline{x},\overline{t}) = \begin{pmatrix}0\\0\end{pmatrix} . \tag{6.14}$$

Here is the connection between (6.13) and the set $C_s(f)$.

Lemma 6.3. *Let $f\in\mathcal{P}_r$. Then $(\overline{x},\overline{t})$ is a critical point of* (6.13) *if and only if $(\overline{x},\overline{t})\in C_s(f)$.*

Proof. The critical point condition (6.14) in particular means that the multiplier $\overline{\lambda}\neq 0$ is a solution of $0=\sum_{i=1}^n \overline{\lambda}_i\nabla_x\frac{\partial}{\partial x_i}f(\overline{x},\overline{t}) = \nabla_x^2 f(\overline{x},\overline{t})\overline{\lambda}$. So, $\det\nabla_x^2 f(\overline{x},\overline{t})=0$ and thus $(\overline{x},\overline{t})\in C_s(f)$.
Suppose now $(\overline{x},\overline{t})\in C_s(f)$, i.e., $\nabla_x^T f(\overline{x},\overline{t})=0$ and $\det\nabla_x^2 f(\overline{x},\overline{t})=0$. So, there is a solution $\xi\neq 0$ of $\nabla_x^2 f(\overline{x},\overline{t})\xi=0$. But since $f\in\mathcal{P}_r$ and thus $f\in\mathcal{P}_r^1$ (cf., (6.12)) the matrix $\nabla\nabla_x^T f(\overline{x},\overline{t})$ has rank n and we can conclude

$$\begin{pmatrix}0\\ \alpha\end{pmatrix} := \Big(\nabla\nabla_x^T f(\overline{x},\overline{t})\Big)^T \xi = \sum_{i=1}^n \xi_i \begin{pmatrix}\nabla_x^T\frac{\partial}{\partial x_i}f(\overline{x},\overline{t})\\ \frac{\partial}{\partial t}\frac{\partial}{\partial x_i}f(\overline{x},\overline{t})\end{pmatrix} \neq \begin{pmatrix}0\\0\end{pmatrix} .$$

Consequently, the number $\alpha=\sum_{i=1}^n \xi_i\frac{\partial}{\partial t}\frac{\partial}{\partial x_i}f(\overline{x},\overline{t})$ is not equal to zero, and by dividing the foregoing relation by α we have found a solution $\overline{\lambda}:=-\xi/\alpha$ of the critical point equation (6.14) for the program (6.13).

□

Now let again $f\in\mathcal{P}_r$ and $(\overline{x},\overline{t})\in C_s(f)$, and recall that by Ex. 6.1 we have rank $\nabla_x^2 f(\overline{x},\overline{t})=n-1$. By applying an appropriate linear transformation $x\to Ux$, wlog. we can

$$\text{assume:}\qquad \nabla_x^2 f(\overline{x},\overline{t}) = \text{diag}\,(0,\sigma_2,\dots,\sigma_n)\ ,\quad \sigma_i\neq 0,\ i=2,\dots,n\ . \tag{6.15}$$

Then since $(\overline{x},\overline{t})\in C_s(f)$ (see Lemma 6.3), $(\overline{x},\overline{t})$ solves the critical point equations (6.14), where the first n equations have the form $\nabla_x^2 f(\overline{x},\overline{t})\overline{\lambda}=(\overline{\lambda}_1 0,\overline{\lambda}_2\sigma_2,\dots,\overline{\lambda}_n\sigma_2)=0$ with $\overline{\lambda}\neq 0$. Consequently we find

$$\overline{\lambda}_1\neq 0,\quad \overline{\lambda}_i=0,\ i=2,\dots,n\ . \tag{6.16}$$

The last equation in (6.14) then implies $-1=\overline{\lambda}_1\frac{\partial}{\partial t}\frac{\partial}{\partial x_1}f(\overline{x},\overline{t})$ or

$$\frac{\partial}{\partial t}\frac{\partial}{\partial x_1}f(\overline{x},\overline{t})\neq 0\ . \tag{6.17}$$

The next lemma characterizes the nondegenerate critical points of (6.13) (see Definition 6.2 below).

Lemma 6.4. *Let $f\in\mathcal{P}_r$ and let $(\overline{x},\overline{t})$ be a critical point of* (6.13)*, i.e., $(\overline{x},\overline{t})\in C_s(f)$. Then (under the assumption* (6.15)*), $(\overline{x},\overline{t})$ is a nondegenerate critical point of* (6.13) *if and* *only if*

$$\frac{\partial^3}{\partial x_1^3} f(\overline{x},\overline{t}) \neq 0 \ . \tag{6.18}$$

Proof. For $f \in \mathcal{P}_r$ in particular we have rank $\nabla\nabla_x^T f(\overline{x},\overline{t}) = n$ if $(\overline{x},\overline{t}) \in C(f)$, where the rows of $\nabla\nabla_x^T f(\overline{x},\overline{t})$ are the gradients of the constraint functions $\frac{\partial}{\partial x_i} f(\overline{x},\overline{t})$, $i = 1,\dots,n$, and are linearly independent. Under (6.15) we find

$$\nabla\nabla_x^T f(\overline{x},\overline{t}) = \left(\begin{array}{c|c} \text{diag}\,(0,\sigma_2,\dots,\sigma_n) & \begin{array}{c} \frac{\partial^2}{\partial t \partial x_1} f(\overline{x},\overline{t}) \\ \vdots \\ \frac{\partial^2}{\partial t \partial x_n} f(\overline{x},\overline{t}) \end{array} \end{array} \right) \ .$$

Obviously, the tangent space

$$T_{(\overline{x},\overline{t})} = \{\xi \in \mathbb{R}^{n+1} \mid \nabla \frac{\partial}{\partial x_i} f(\overline{x},\overline{t})\xi = 0,\ i = 1,\dots,n\} = \ker \nabla\nabla_x^T f(\overline{x},\overline{t})$$

is spanned by $\xi = (1,0,\dots,0) \in \mathbb{R}^{n+1}$. By Definition 6.2 below, the critical point $(\overline{x},\overline{t})$ of (6.13) is nondegenerate if with $\overline{\lambda}$ in (6.16) it holds

$$0 \neq \xi^T \nabla^2 L(\overline{x},\overline{t},\overline{\lambda})\xi = \overline{\lambda}_1 \xi^T \nabla^2 \frac{\partial}{\partial x_1} f(\overline{x},\overline{t})\xi = \overline{\lambda}_1 \frac{\partial^3}{\partial x_1^3} f(\overline{x},\overline{t}) \ .$$

In view of $\overline{\lambda}_1 \neq 0$ this is equivalent with $\frac{\partial^3}{\partial x_1^3} f(\overline{x},\overline{t}) \neq 0$. □

Recall that for $f \in \mathcal{P}_r$ the solution set $C(f)$ of $\nabla_x^T f(x,t) = 0$ is a one-dimensional C^∞-manifold of $\mathbb{R}^n \times \mathbb{R}$. Let us examine this manifold near a point $(\overline{x},\overline{t}) \in C_s(f)$. We again assume (6.15). Then ($\sigma_i \neq 0,\ i = 2,\dots,n$)

$$\nabla\nabla_x^T f(\overline{x},\overline{t}) = \left(\text{diag}\,(0,\sigma_2,\dots,\sigma_n) \,\middle|\, \frac{\partial}{\partial t}\nabla_x^T f(\overline{x},\overline{t}) \right) \quad \text{has rank } \ n \ ,$$

and also the $(n \times n)$-submatrix (with $x^2 := (x_2,\dots,x_n)$, $f_{it} := \frac{\partial^2}{\partial x_i \partial t} f(\overline{x},\overline{t})$)

$$\nabla_{(x^2,t)} \nabla_x^T f(\overline{x},\overline{t}) = \left(\begin{array}{ccc|c} 0 & \cdots & 0 & f_{1t} \\ \sigma_2 & \cdots & 0 & f_{2t} \\ & \ddots & & \vdots \\ 0 & \cdots & \sigma_n & f_{nt} \end{array} \right) \quad \text{has rank } \ n \ . \tag{6.19}$$

The Implicit Function Theorem applied to $\nabla_x^T f(x,t) = 0$ assures that near $(\overline{x},\overline{t})$ the set $C(f)$ can be parameterized via a C^∞-function $(x^2(x_1), t(x_1)),\ x_1 \approx \overline{x}_1$, such that

$$\nabla_x^T f(x_1, x^2(x_1), t(x_1)) \equiv 0 \quad \text{for } x_1 \text{ near } \overline{x}_1 \ . \tag{6.20}$$

Differentiation wrt. x_1, using $\nabla_x^T \frac{\partial}{\partial x_1} f(\overline{x},\overline{t}) = 0$ (cf., (6.15)) yields

$$\nabla_{(x^2,t)} \nabla_x^T f(\overline{x},\overline{t}) \begin{pmatrix} \frac{d}{dx_1} x^2(\overline{x}_1) \\ \frac{d}{dx_1} t(\overline{x}_1) \end{pmatrix} = -\nabla_x^T \frac{\partial}{\partial x_1} f(\overline{x},\overline{t}) = 0 \ . \tag{6.21}$$

So in view of (6.19) we obtain

$$\frac{d}{dx_1}x^2(\overline{x}_1) = 0 \; , \quad \frac{d}{dx_1}t(\overline{x}_1) = 0 \; . \tag{6.22}$$

The first row of relation (6.21) reads

$$\nabla_{(x^2,t)}\frac{\partial}{\partial x_1}f(\overline{x},\overline{t})\begin{pmatrix} \frac{d}{dx_1}x^2(\overline{x}_1) \\ \frac{d}{dx_1}t(\overline{x}_1) \end{pmatrix} = -\frac{\partial^2}{\partial x_1^2}f(\overline{x},\overline{t})$$

and a further differentiation wrt. x_1 leads to (recall $\left(\frac{d}{dx_1}x^2(\overline{x}_1), \frac{d}{dx_1}t(\overline{x}_1)\right)^T = 0$)

$$0 + \nabla_{(x^2,t)}\frac{\partial}{\partial x_1}f(\overline{x},\overline{t})\begin{pmatrix} \frac{d^2}{dx_1^2}x^2(\overline{x}_1) \\ \frac{d^2}{dx_1^2}t(\overline{x}_1) \end{pmatrix} = -\frac{\partial^3}{\partial x_1^3}f(\overline{x},\overline{t}) \; .$$

In view of $\nabla_{(x^2,t)}\frac{\partial}{\partial x_1}f(\overline{x},\overline{t}) = \left(0, \frac{\partial^2}{\partial t \partial x_1}f(\overline{x},\overline{t})\right)$ it follows

$$\frac{\partial^2}{\partial t \partial x_1}f(\overline{x},\overline{t})\;\frac{d^2}{dx_1^2}t(\overline{x}_1) = -\frac{\partial^3}{\partial x_1^3}f(\overline{x},\overline{t}) \quad \text{or} \quad \frac{d^2}{dx_1^2}t(\overline{x}_1) = \frac{-\frac{\partial^3}{\partial x_1^3}f(\overline{x},\overline{t})}{\frac{\partial^2}{\partial t \partial x_1}f(\overline{x},\overline{t})} \; . \tag{6.23}$$

Recall that by (6.17) we have $\frac{\partial^2}{\partial t \partial x_1}f(\overline{x},\overline{t}) \neq 0$. So, under the assumption $\frac{\partial^3}{\partial x_1^3}f(\overline{x},\overline{t}) \neq 0$, for the function $t(x_1)$ we find (see (6.22)):

$$\frac{d}{dx_1}t(\overline{x}_1) = 0 \quad \text{and} \quad \frac{d^2}{dx_1^2}t(\overline{x}_1) \neq 0 \; . \tag{6.24}$$

Consequently, in this case, the program (6.13) has on $C(f)$ (see Figure 6.9)

$$\text{a strict local maximizer (or minimizer) at } (\overline{x},\overline{t}) \text{ if } \frac{d^2}{dx_1^2}t(\overline{x}_1) < 0 \;\; (\text{or } > 0) \; . \tag{6.25}$$

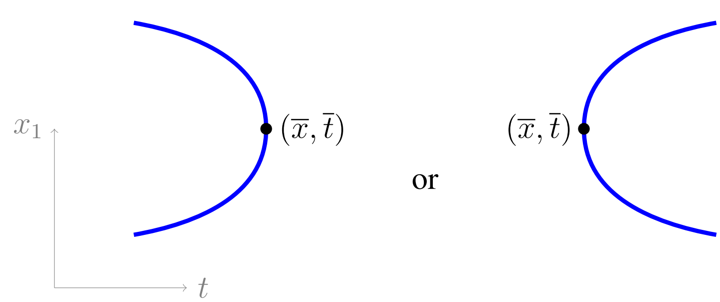

FIGURE 6.9. Sketch of $C(f)$ with a turning point $(\overline{x},\overline{t})$.

In other words, near $(\overline{x},\overline{t}) \in C_s(f)$, the curve $C(f)$ shows a "quadratic turning point". The following theorem states that in the generic case, i.e., for $f \in \mathcal{P}_r$, at all points $(\overline{x},\overline{t}) \in C_s(f)$ such quadratic turning points occur.

THEOREM 6.7. *Let $f \in \mathcal{P}_r$. Then for all $(\overline{x}, \overline{t}) \in C_s(f)$ the following holds.*
(a) All critical points $(\overline{x}, \overline{t})$ of (6.13) *(i.e., $(\overline{x}, \overline{t}) \in C_s(f)$, see Lemma 6.3) are nondegenerate critical points of* (6.13).
In particular by Lemma 6.4 and (6.24)*, around all points $(\overline{x}, \overline{t}) \in C_s(f)$ the set $C(f)$ has a quadratic turning point.*

(b) Let under assumption (6.15) *near $(\overline{x}, \overline{t}) \in C_s(f)$ the solution set $C(f)$ be parameterized by $(x_1, x^2(x_1), t(x_1)),\ x_1 \approx \overline{x}_1$ (see* (6.20)*). Then for the determinant of $\nabla_x^2 f(x,t)$ it holds*

$$\frac{d}{dx_1} \det f(\overline{x}_1, x^2(\overline{x}_1), t(\overline{x}_1)) \neq 0 \ .$$

This means (together with $\det \nabla_x^2 f(\overline{x}, \overline{t}) = 0$, see (6.10)*) that the determinant $\det \nabla_x^2 f(x,t)$ changes sign when passing on the curve $C(f)$ a point $(\overline{x}, \overline{t}) \in C_s(f)$. Moreover, the number of positive (negative) eigenvalues of $\nabla_x^2 f(x,t)$ changes exactly by one when passing $(\overline{x}, \overline{t}) \in C_s(f)$.*

Proof. For the proofs of (a) and (b) we refer to [21, Theorem 10.23], but for completeness we reproduce the proof of (b).
(b) Under assumption (6.15), by applying the multi-linearity of the determinant, it follows with the abreviations $f_{ij} := \frac{\partial^2}{\partial x_i \partial x_j} f(\overline{x}, \overline{t})$, $f_{111} := \frac{\partial^3}{\partial x_1^3} f(\overline{x}, \overline{t})$ etc., and using $(f_{11}, \ldots, f_{n,1}) = (0, \ldots, 0)$ (see (6.15)):

$$\begin{aligned}
\frac{\partial}{\partial x_1} \det \nabla_x^2 f(\overline{x}, \overline{t}) &= \frac{\partial}{\partial x_1} \det \begin{pmatrix} f_{11} & \cdots & f_{1i} & \cdots & f_{1n} \\ \vdots & & \vdots & & \vdots \\ f_{n1} & \cdots & f_{ni} & \cdots & f_{nn} \end{pmatrix} \\
&= \sum_{i=1}^{n} \det \begin{pmatrix} f_{11} & \cdots & f_{11i} & \cdots & f_{1n} \\ \vdots & & \vdots & & \vdots \\ f_{n1} & \cdots & f_{1ni} & \cdots & f_{nn} \end{pmatrix} \\
&= \det \begin{pmatrix} f_{111} & 0 & \cdots & 0 \\ f_{112} & \sigma_2 & \cdots & 0 \\ \vdots & & \ddots & 0 \\ f_{11n} & 0 & \cdots & \sigma_n \end{pmatrix} \\
&= f_{111} \cdot \sigma_2 \cdots \sigma_n \ .
\end{aligned}$$

Recalling $\sigma_i \neq 0,\ i = 2, \ldots, n$, by applying (a) combined with Lemma 6.4 it follows $\frac{\partial}{\partial x_1} \det \nabla_x^2 f(\overline{x}, \overline{t}) \neq 0$.
It is well-known that the eigenvalues depend continuously on the matrix elements. So near $(\overline{x}, \overline{t})$ the $n-1$ eigenvalues of $\nabla_x^2 f(x,t)$ near $\sigma_2, \ldots, \sigma_n$ will not change sign, and in view of $\frac{\partial}{\partial x_1} \det \nabla_x^2 f(\overline{x}, \overline{t}) \neq 0$ only the eigenvalue near $\sigma_1 = 0$ will change from positive to negative (or from negative to positiv).

□

We finish the section with a remark on the set

$$C_m(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \text{ is a local minimizer of } P(t)\} \ ,$$

corresponding to local minimizers of $P(t)$ (cf., (6.4)) for the one-parametric case $t \in \mathbb{R}$. Clearly this set $C_m(f)$ is a subset of $C(f)$. It is also evident that for functions $f \in \mathcal{P}_r$, on a connected component of the one-dimensional manifold $C(f) \setminus C_s(f)$ the (n nonzero) eigenvalues of $\nabla_x^2 f(x,t)$ do not change signs. So, supposing that on one "side" of the set $C(f)$ near a point $(\overline{x}, \overline{t}) \in C_s(f)$ the elements $(x,t) \in C(f)$ are such that the points x are strict local minimizers of $P(t)$ (i.e., all eigenvalues of $\nabla_x^2 f(x,t)$ are positive). Then according to Theorem 6.7(b) (see Figure 6.9) at the other "side" the Hessian $\nabla_x^2 f(x,t)$ has exactly one negative eigenvalue ($n-1$ eigenvalues will remain positive) and these points $(x,t) \in C(f)$ cannot correspond to local minimizers. In case of $x \in \mathbb{R}$ (only one eigenvalue) on this side local maximizers occur. Consequently, generically, the set $C_m(f)$ looks as sketched in Figure 6.10. We also emphasize that for $f \in \mathcal{P}_r$, because of the

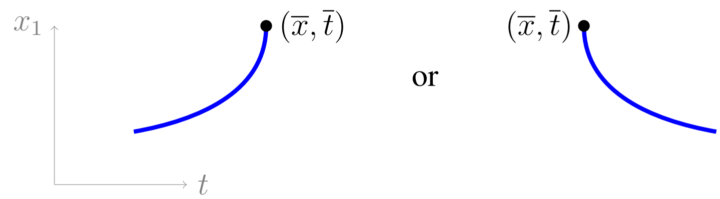

FIGURE 6.10. Sketch of $C_m(f)$ near a point $(\overline{x}, \overline{t}) \in C_s(f)$, for $f \in \mathcal{P}_r$.

conditions $\frac{\partial}{\partial x_1} f(\overline{x}, \overline{t}) = \frac{\partial^2}{\partial x_1^2} f(\overline{x}, \overline{t}) = 0$ and $\frac{\partial^3}{\partial x_1^3} f(\overline{x}, \overline{t}) \neq 0$ (under (6.15)), the critical point $(\overline{x}, \overline{t}) \in C_s(f)$ can never yield a minimizer (or a maximizer) $\overline{x}$ of $P(\overline{t})$, i.e., this point does not belong to $C_m(f)$.

6.4. Genericity results for parametric systems of nonlinear equations

In this section we examine systems of equations depending on a parameter $t \in \mathbb{R}^p$,

$$E_0(t): \qquad F(x,t) = 0 \,, \qquad \text{where } F \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}^m) \,.$$

For any fixed $\overline{t} \in \mathbb{R}^p$ the problem $E_0(\overline{t})$ consists of a system of equations as discussed in Section 4.4. Assume that the following condition holds:

(6.26) $\quad \nabla_{(x,t)} F(x,t)$ has full rank m at all solutions $(x,t) \in \mathbb{R}^n \times \mathbb{R}^p$ of $F(x,t) = 0$.

Under this condition the solution set

$$S(F) := \{(x,t) \in \mathbb{R}^n \times \mathbb{R}^p \mid F(x,t) = 0\}$$

is a C^∞-manifold of codimension m (dimension $n+p-m$) in $\mathbb{R}^n \times \mathbb{R}^p$. Indeed under (6.26) the set $S(F)$ satisfies the condition CM1 (cf., Section 5.1).

As main result of this section we will show that the condition (6.26) is a generic property of $E_0(t)$.
Before analyzing the generic structure of $E_0(t)$ we give some illustrative examples.

<u>$n = p = m = 1$</u>: $F(x,t) := x^2 - t^2 = 0$ with (see Figure 6.11)

$$S(F) = \{(x,t) \in \mathbb{R} \times \mathbb{R} \mid x^2 - t^2 = 0\} .$$

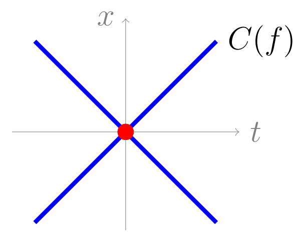

FIGURE 6.11. Solution set $S(F)$ for $F(x,t) := x^2 - t^2 = 0$.

Note, that at the solution point $(\overline{x}, \overline{t}) = (0,0) \in S(F)$ the "regularity condition" in (6.26) is not fulfilled: rank $\nabla_{(x,t)} F(\overline{x}, \overline{t}) =$ rank $(0,0) = 0 < m$.
For the perturbed equation $F_\pm(x,t) := x^2 - t^2 \pm \varepsilon = 0$ ($\varepsilon > 0$ fixed) the sets $S(F_\pm)$ are sketched in Figure 6.12.

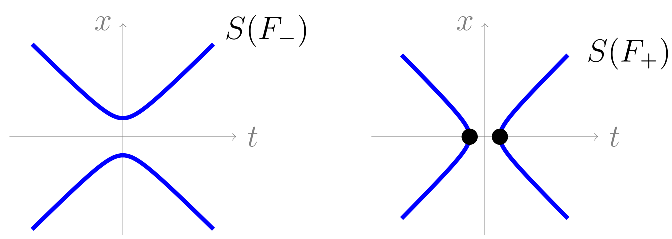

FIGURE 6.12. Solution sets $S(F_\pm)$ of the perturbed equations $x^2 - t^2 \pm \varepsilon = 0$.

Both sets $S(F_\pm)$ are manifolds of dimension 1. For F_- even the stronger condition is satisfied:

$$\nabla_x F(x,t) \text{ has full rank } m \text{ at all solutions } (x,t) \in \mathbb{R}^n \times \mathbb{R}^p \text{ of } F(x,t) = 0 .$$

So by the Implicit Function Theorem 7.2 the set $S(F_-)$ can be parameterized by functions $x(t)$. This is in contrast to $S(F_+)$ where at the solution points $(0, \pm\sqrt{\varepsilon}) \in S(F_+)$ we have rank $\nabla_{(x,t)} F(\overline{x}, \overline{t}) =$ rank $(0, \pm 2\sqrt{\varepsilon}) = 1 = m$ but rank $\nabla_x F(\overline{x}, \overline{t}) =$ rank $0 = 0 < m$, and $S(F_+)$ has so-called turning points at $(0, \pm\sqrt{\varepsilon})$.

<u>$n = p = 1,\ m = 2$</u>: $F(x,t) := (x - t - 1, x - t^2) = (0,0)$. Obviously the solution set

$$S(F) = \{(x,t) \in \mathbb{R} \times \mathbb{R} \mid x = t + 1, \;\; x = t^2\} = \left\{ \frac{1}{2}(1 \pm \sqrt{5}, 3 \pm \sqrt{5}) \right\}$$

consists of two points (cf. Figure 6.13).
Also here the condition (6.26) is valid and $S(F)$ is a nice manifold, however in this case , of dimension 0 in $\mathbb{R}^2$.

We will now show, that the condition (6.26) is generically fulfilled.

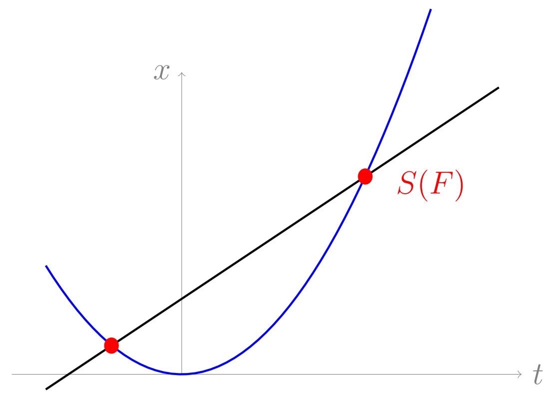

FIGURE 6.13. Set $S(F)$ for $F(x,t) = (x - t - 1, x - t^2) = (0,0)$ (red points).

THEOREM 6.8. *The set*

$$\mathcal{P}_r = \{F \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}^m) \mid F \text{ satisfies } (6.26)\ \} . \tag{6.27}$$

is open and dense (thus generic) in $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}^m)$ wrt. the C^k_s-topology, $k \geq 1$. Moreover, for $F \in \mathcal{P}_r$ the solution set $S(F)$ is a manifold in $\mathbb{R}^n \times \mathbb{R}^p$ of codimension m.

Proof. Consider the function $j^0 F = \text{graph } (F)$ and the reduced jet-function $\tilde{j}^0 F(x,t) = F(x,t)$. By Theorem 5.15 the condition $F \pitchfork \{0\}$ holds iff for any solution (x,t) of $F(x,t) = 0$ we have (use $T_{F(x,t)}\{0\} = \{0\}$)

$$\nabla F(x,t)[\mathbb{R}^n \times \mathbb{R}^p] + \{0\} = \mathbb{R}^m ,$$

or equivalently (6.26). Consequently, $F \pitchfork \{0\}$ iff $F \in \mathcal{P}_r$. According to Lemma 6.1 $F \pitchfork \{0\}$ holds iff $j^0 F \pitchfork (\mathbb{R}^n \times \mathbb{R}^p \times \{0\})$, and since $(\mathbb{R}^n \times \mathbb{R}^p \times \{0\})$ is closed, the statement follows from the Jet Transversality Theorem 6.1.
The fact that for $f \in P_r$ the set $S(f)$ is a manifold of codimension m has been shown before. □

We just present a special case.

- case $n + p < m$: This in particular implies $n + p <$ codim $\{0\}, 0 \in \mathbb{R}^m$, and by Lemma 5.6 (with n replaced by $n+p$) it follows that generically, i.e., for $F \in \mathcal{P}_r$, we have $\{F(x,t) \mid (x,t) \in \mathbb{R}^n \times \mathbb{R}^p\} \cap \{0\} = \emptyset$. In other words we have

$$S(F) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R}^p \mid F(x,t) = 0\} = \emptyset .$$

Let us again discuss a stronger condition than in (6.26) which is related with the condition LICQ (see Section 6.5, (6.31)). Considering the system

$$F(x,t) = \big(g_1(x,t), \dots, g_m(x,t)\big)^T = 0, \quad g_j \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}),\ j = 1, \dots, m ,$$

we say that at a solution $(\overline{x}, \overline{t}) \in \mathbb{R}^n \times \mathbb{R}^p$ the condition LICQ is satisfied if the gradients $\nabla_x g_j(\overline{x}, \overline{t})$, $j = 1, \dots, m$, are linearly independent or equivalently if rank $\nabla_x F(\overline{x}, \overline{t}) = m$.

So, a nice condition would be the following (cf. (6.26)):

(6.28) $\quad \nabla_x F(x,t)$ has full rank m at all solutions $(x,t) \in \mathbb{R}^n \times \mathbb{R}^p$ of $F(x,t) = 0$.

We can proceed as before in Subsection 6.3.1. By taking the Whitney regular stratification of the closed set $\{0\} \times M^{m,n}$ given by

$$\big(\{0\} \times M^{m,n}, \hat{\Sigma}\big) \quad \text{with} \quad \hat{\Sigma} = \big\{\hat{M}_j := \{0\} \times M^{m,n}(j),\ j = 0,\ldots,\min\{n,m\}\big\}\ ,$$

and by defining the reduced 1-jet function

$$\hat{j}^1 F(x,t) = \big(F(x,t), \nabla_x F(x,t)\big)$$

we obtain the following theorem for the set

$$\hat{\mathcal{P}}_r = \{F \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}^m) \mid \hat{j}^1 F(x,t) \pitchfork \hat{M}_j,\ \forall \hat{M}_j \in \hat{\Sigma}\}\ . \tag{6.29}$$

THEOREM 6.9. *The set $\hat{\mathcal{P}}_r$ is open and dense (thus generic) in $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R}^m)$ wrt. the C^k_s-topology, $k \geq 2$. Moreover, for $F \in \hat{\mathcal{P}}_r$ the inverse images*

$$(\hat{j}^1 F)^{-1}\big(\{0\} \times M^{m,n}(j)\big)\ , \quad j = 0,\ldots,\min\{n,m\}\ ,$$

are submanifolds in $S(F)$ of codimension $c_d = \text{codim}\,\big(M^{m,n}(j)\big) = (m-j)(n-j)$, and these sets form a Whitney regular stratification of the set $S(F) = (\hat{j}^1 F)^{-1}\big(\{0\} \times M^{m,n}\big)$.

Let us also consider some special cases.

- case $n + p < m$: Then as before, for $F \in \hat{\mathcal{P}}_r$ we find
$$S(F) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R}^p \mid F(x,t) = 0\} = \emptyset\ .$$
- case $n + p = m$: In particular $m \geq n$. By Theorem 6.9 the sets $(\hat{j}^1 F)^{-1}(\{0\} \times M^{m,n}(j))$ are manifolds of codimension $c_d = (m-j)(n-j)$ in $S(F)$. So for $F \in \hat{\mathcal{P}}_r$,

$$\big\{(x,t) \in \mathbb{R}^{n+p} \mid F(x,t) = 0,\ \text{rank}\ \nabla_x F(x,t) = j\big\} = \begin{cases} \text{discrete set} & \text{for } j = n \\ \emptyset & \text{for } j < n \end{cases}\ .$$

- general case for $F \in \hat{\mathcal{P}}_r$: If $n + p < m + (m-j)(n-j)$ then
$$\big\{(x,t) \in \mathbb{R}^{n+p} \mid F(x,t) = 0,\ \text{rank}\ \nabla_x F(x,t) = j\big\} = \emptyset\ .$$

6.5. Genericity results for parametric constrained programs

This section deals with parametric constrained programs of the form: For $t \in \mathbb{R}^p$ solve

$$P(t): \quad \min\ f(x,t) \quad \text{s.t.} \quad x \in \mathcal{F}(t) = \{x \in \mathbb{R}^n \mid g_j(x,t) \leq 0,\ j \in J\}\ , \tag{6.30}$$

where $J = \{1,\ldots,m\}$ and the functions $f, g_j, j \in J$, are in $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R})$. For fixed $\bar{t} \in \mathbb{R}^p$ the problem $P(\bar{t})$ represents a constrained program as studied in Section 4.5.

We again use the abbreviation $G = (g_j,\ j \in J)^T$ and often write $P(f,G)$, $\mathcal{F}(G)$ instead of $P(t)$, $\mathcal{F}(t)$, if we wish to stress on the fact that the program is defined by the problem functions (f,G).

Let $(\overline{x}, \overline{t})$, $\overline{x} \in \mathcal{F}(\overline{t}),\ \overline{t} \in \mathbb{R}^p$. As before we define the active index set $J_{(\overline{x},\overline{t})} = \{j \in J \mid g_j(\overline{x}, \overline{t}) = 0\}$ and the Lagrangean function near $(\overline{x}, \overline{t})$:

$$L(x, t, \mu) = f(x, t) + \sum_{j \in J_{(\overline{x},\overline{t})}} \mu_j g_j(x, t) \ .$$

Recall that the linear independence constrained qualification (LICQ) is said to hold at $(\overline{x}, \overline{t})$ if the gradients

(6.31) $$\nabla_x g_j(\overline{x}, \overline{t}), \quad j \in J_{(\overline{x},\overline{t})} \ , \qquad \text{are linearly independent} \ .$$

For a better understanding of the structure of the parametric problem $P(t)$ we will not only look at local minimizers of $P(t)$ but also at generalized critical points.

Definition 6.1. *Given a problem $P(t) = P(f, G)$, let $(\overline{x}, \overline{t}),\ \overline{x} \in \mathcal{F}(\overline{t}),\ \overline{t} \in \mathbb{R}^p$.*
(a) The point $(\overline{x}, \overline{t})$ with $\overline{x} \in \mathcal{F}(\overline{t})$ is called a generalized critical point (gc-point) of $P(t)$, if there exist Lagrangean multipliers $\overline{\mu}_0 \in \mathbb{R},\ \overline{\mu}_j \in \mathbb{R},\ j \in J_{(\overline{x},\overline{t})}$, not all zero, such that

(6.32) $$\overline{\mu}_0 \nabla_x f(\overline{x}, \overline{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \overline{\mu}_j \nabla_x g_j(\overline{x}, \overline{t}) = 0 \ ,$$

or equivalently that $\nabla_x f(\overline{x}, \overline{t}), \nabla_x g_j(\overline{x}, \overline{t}),\ j \in J_{(\overline{x},\overline{t})}$, are linearly dependent.
(b) The point $(\overline{x}, \overline{t})$ with $\overline{x} \in \mathcal{F}(\overline{t})$ is said to be a critical point of $P(t)$, if LICQ holds at $(\overline{x}, \overline{t})$, and there are multipliers $\overline{\mu}_j,\ j \in J_{(\overline{x},\overline{t})}$ (unique by LICQ), such that

(6.33) $$\nabla_x f(\overline{x}, \overline{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \overline{\mu}_j \nabla_x g_j(\overline{x}, \overline{t}) = 0 \ .$$

In the sequel C_{gc} denotes the set of all gc-points of $P(t)$ and C_m denotes the set $C_m = \{(x, t) \mid x$ is a local minimizer of $P(t)\}$.
Note that (by the John-condition, cf., [13]) any local minimizer of $P(t)$ must necessarily be a gc-point, i.e., $C_m \subset C_{gc}$.

Definition 6.2. *A generalized critical point $(\overline{x}, \overline{t})$ of $P(t)$ is called a nondegenerate critical point if* (6.33) *holds together with*

- *LICQ (i.e., $(\overline{x}, \overline{t})$ is a critical point)*
- *strict complementarity,*

(6.34) $$(SC): \qquad \overline{\mu}_j \neq 0 \ \ \forall j \in J_{(\overline{x},\overline{t})} \ ,$$

- *and the second order condition*

(6.35) $$(SOC): \quad d^T \nabla_x^2 L(\overline{x}, \overline{t}, \overline{\mu}) d \neq 0 \quad \forall d \in T_{(\overline{x},\overline{t})} \setminus \{0\} \ ,$$

where $T_{(\overline{x},\overline{t})}$ is the tangent space $T_{(\overline{x},\overline{t})} = \{d \in \mathbb{R}^n \mid \nabla_x g_j(\overline{x}, \overline{t}) d = 0,\ j \in J_{(\overline{x},\overline{t})}\}$.
The set of nondegenerate critical points is denoted by C_r.

6.5.1. Parametric constrained programs.

In this subsection we study the generic properties of general parametric problems $P(t)$ in (6.30). Recall from Section 4.5 that in the nonparametric case, generically all local minimizers $\overline{x}$ are nondegenerate, i.e., according to Theorem 4.8 the conditions LICQ, SC, and SOC are satisfied at $\overline{x}$. In the same way we could have proven that generically, in the nonparametric case, all gc-points are nondegenerate critical points (cf., Definition 6.2). However again, it is clear that for parametric programs we cannot expect that generically for all $t \in \mathbb{R}^p$ all gc-points of $P(t)$ are nondegenerate critical points. More precisely, the higher the dimension p of the parameter space $\mathbb{R}^p$ is, the more degeneracies may occur in the generic situation.

In what follows, we are interested in a more detailed description of the generic structure of the set C_r of gc-points $(x,t) \in \mathbb{R}^n \times \mathbb{R}^p$ which are nondegenerate and the set $C_s := C_{gc} \backslash C_r$, where at least one of the nice conditions LICQ, SC, or SOC fails.

We start with the set C_r of a given problem $P(t) = P(f,G)$. A critical point $(\overline{x},\overline{t})$ is a solution of the system,

$$
H(x,t,\mu) := \begin{array}{rl} \nabla_x^T f(x,t) + \sum_{j \in J_{(\overline{x},\overline{t})}} \ \mu_j \nabla_x^T g_j(x,t) & = 0 \\ g_j(x,t) & = 0 \, , \quad j \in J_{(\overline{x},\overline{t})} \end{array} \tag{6.36}
$$

with some $\overline{\mu}$. If $(\overline{x},\overline{t})$ is nondegenerate, then according to Ex. 4.4 the Jacobian $H_{(x,\mu)}$ of H at $(\overline{x},\overline{t},\overline{\mu})$ ($G_{J_{(\overline{x},\overline{t})}} = (g_j(\overline{x},\overline{t}),\ j \in J_{(\overline{x},\overline{t})})^T$),

$$
H_{(x,\mu)}(\overline{x},\overline{t},\overline{\mu}) = \begin{pmatrix} \nabla_x^2 L(\overline{x},\overline{t},\overline{\mu}) & \nabla_x G_{J_{(\overline{x},\overline{t})}}(\overline{x},\overline{t})^T \\ \nabla_x G_{J_{(\overline{x},\overline{t})}}(\overline{x},\overline{t}) & 0 \end{pmatrix} , \quad \text{is nonsingular.}
$$

So near $(\overline{x},\overline{t},\overline{\mu})$ we can apply the Implicit Function Theorem 7.2 to the system (6.36). This theorem implies that in a neighborhood $U_{(\overline{x},\overline{t})} \times U_{\overline{\mu}}$ of $(\overline{x},\overline{t},\overline{\mu})$ the solution set of (6.36) is a p-dimensional manifold parameterized by locally defined C^∞-functions $(x(t),\mu(t))$. Moreover, the points $(x(t),t) \in U_{(\overline{x},\overline{t})}$ are nondegenerate critical points of $P(t)$, see Figure 6.14.

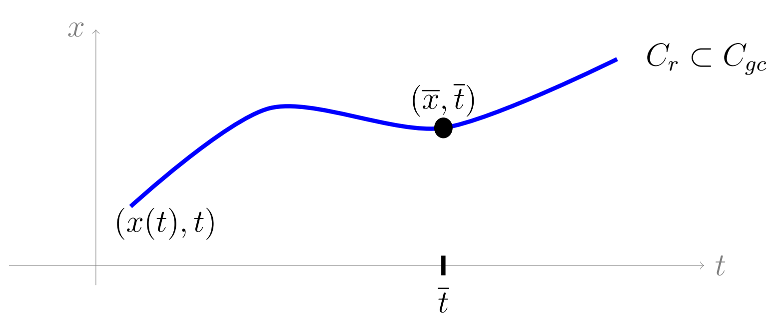

FIGURE 6.14. Local structure of C_r near a nondegenerate critical point $(\overline{x},\overline{t})$ of $P(t)$.

We now look at the set $C_s = C_{gc} \setminus C_r$ of gc-points where at least one of the conditions

LICQ, SC, or SOC fails. We give a rough picture for the general case of parametric programs $P(t)$, $t \in \mathbb{R}^p$. Only for the one-parametric case ($p = 1$) a complete description has been accomplished until now (see Subsection 6.5.2 for details).

We are interested in points $(x,t) \in \mathbb{R}^n \times \mathbb{R}^p$ from the singular critical set C_s and consider the failure of LICQ, SC, and SOC for a program $P(t) = P(f,G)$. To that end we introduce the reduced 2-jet function $\tilde{j}^2(f,G) : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m \times \big(\mathbb{R}^n\big)^{m+1} \times \big(S^n\big)^{m+1}$ defined by

(6.37)
$$\tilde{j}^2(f,G)(x,t) := \big(G(x,t), \nabla_x f(x,t), \nabla_x g_j(x,t),\ j \in J, \nabla_x^2 f(x,t), \nabla_x^2 g_j(x,t),\ j \in J\big)\ .$$

In what follows, we identify ($G = (g_j,\ j \in J)^T$)

$$a \equiv G(x,t),\quad b_0 = \nabla_x^T f(x,t),\quad B \equiv \nabla_x G(x,t)^T,$$
$$A_0 \equiv \nabla_x^2 f(x,t),\quad A_j \equiv \nabla_x^2 g_j(x,t),\ j \in J\ , \tag{6.38}$$

and translate the failure of LICQ , SC, SOC into semi-algebraic conditions in the variables

$$\big(a, b_0, B, A_0, A_j,\ j \in J\big) \in \mathbb{R}^m \times \big(\mathbb{R}^n\big)^{m+1} \times \big(S^n\big)^{m+1} =: \mathcal{N}\ . \tag{6.39}$$

LICQ: Points $(x,t) \in C_s$, where LICQ is not fulfilled with $J_{(x,t)} = J_0$ for some $J_0 \subset J$, satisfy the conditions

$$\begin{aligned} G_{J_0}(x,t) &= 0 \\ \text{rank}\,\big(\nabla_x G_{J_0}(x,t)\big) &= k < |J_0| \le m\ . \end{aligned}$$

In terms of the variables $a_{J_0}, B_{J_0} = [B_{\cdot j},\ j \in J_0]$ (see (6.38)) these conditions (for J_0 fixed) read,

$$a_{J_0} = 0,\quad \text{rank}\ B_{J_0} = k < |J_0| \le m\ . \tag{6.40}$$

They can easily be shown to be linearly independent and they define a C^∞-semi-algebraic manifold $S_{J_0}(k) \subset \mathcal{N}$ (see (6.39)) of codimension (cf., Theorem 5.11),

$$\text{codim}\ S_{J_0}(k) = |J_0| + (n-k)(|J_0|-k) \ge |J_0| + (n - |J_0| + 1) = n+1\ .$$

Note that in case LICQ fails at (x,t), this point is automatically a gc-point. Moreover, for $|J_0| > n+p$, generically, there is no solution of $a_{J_0} \equiv G_{J_0}(x,t) = 0$ (cf., the special case "$n + p < m$" at the end of Section 6.3).

SC: We consider points $(x,t) \in C_s$, where SC fails but LICQ holds, with $J_{(x,t)} = J_0$ for some $J_0 \subset J$. Such points satisfy for some $j_0 \in J_0$ (with $\mu_{j_0} = 0$) the following conditions:

$$\begin{aligned} G_{J_0}(x,t) &= 0 \\ \text{rank}\left(\nabla_x G_{J_0\setminus\{j_0\}}(x,t)^T \mid \nabla_x^T f(x,t)\right) &= k \le |J_0| - 1\ . \end{aligned}$$

In terms of the variables $a_{J_0}, B_{J_0\setminus\{j_0\}}, b_0$ (see (6.38)) these relations give,

$$a_{J_0} = 0,\quad \text{rank}\ \big(B^T_{J_0\setminus\{j_0\}} \mid b_0\big) = k \le |J_0| - 1\ , \tag{6.41}$$

and these conditions are linearly independent. They also define a C^∞-semi-algebraic manifold $S_{J_0,j_0}(k) \subset \mathcal{N}$ (see (6.39)) of codimension (cf., Theorem 5.11),

$$\text{codim } S_{J_0,j_0}(k) = |J_0| + (n-k)(|J_0| - k) \geq |J_0| + (n - |J_0| + 1) = n + 1 \; .$$

<u>SOC:</u> Here, we look at points $(x,t) \in C_s$, where SOC fails but LICQ holds, with $J_{(x,t)} = J_0$ for some $J_0 \subset J$. Such a point in particular satisfies the conditions, rank $\big(\nabla_x G_{J_0}(x,t)\big) = |J_0|$ and

$$(6.42) \qquad G_{J_0}(x,t) = 0, \quad \text{rank} \left(\nabla_x G_{J_0}(x,t)^T \mid \nabla_x^T f(x,t) \right) = |J_0| \leq n \; .$$

We again identify $a_{J_0} \equiv G_{J_0}(x,t),\ B_{J_0} \equiv \nabla_x G_{J_0}(x,t)^T,\ b_0 \equiv \nabla_x^T f(x,t)$ etc. Then the critical point equation (6.33) yields $B_{J_0}\mu = -b_0$ or after multiplication by $B_{J_0}^T$, $B_{J_0}^T B_{J_0}\mu = -B_{J_0}^T b_0$ and since $B_{J_0}^T B_{J_0}$ is nonsingular the unique Lagrangean multiplier is given by

$$\mu = \mu(b_0, B_{J_0}) = -\big(B_{J_0}^T B_{J_0}\big)^{-1} B_{J_0}^T b_0 \; .$$

The failure of SOC (see Ex. 4.4) means that with the matrix $A := A_0 + \sum_{j \in J_0} \mu_j A_j \in S^n$ (recall $A_0 \equiv \nabla_x^2 f(x,t)$ etc.), the relation

$$(6.43) \qquad \text{rank} \begin{pmatrix} A & B_{J_0} \\ B_{J_0}^T & 0 \end{pmatrix} = k \leq n + |J_0| - 1 \; ,$$

holds. Again, the conditions (6.42) and (6.43) (together with LICQ, i,e, rank $B_{J_0} = |J_0|$) define a semi-algebraic set $S_{J_0}^{SOC}(k) \subset \mathcal{N}$ (cf., (6.39)) of codimension (in both cases $|J_0| < n$ and $|J_0| = n$)

$$\text{codim } S_{J_0}^{SOC}(k) \geq |J_0| + (n - |J_0|)(|J_0| + 1 - |J_0|) + (n + |J_0| - n - |J_0| + 1)^2 = n + 1 \; .$$

Now we can consider the union of all semi-algebraic sets $S_{J_0}(k), S_{J_0,j_0}(k), S_{J_0}^{SOC}(k)$ of codimension $\geq n+1$ above, corresponding to the sets where LICQ, SC, or SOC are not satisfied:

$$S_1 := \bigcup_{\substack{k \leq |J_0|-1, \\ J_0 \subset J}} S_{J_0}(k) \; , \quad S_2 = \bigcup_{\substack{k \leq |J_0|-1, \\ J_0 \subset J, j_0 \in J_0}} S_{J_0,j_0}(k) \; , \quad S_3 = \bigcup_{\substack{k \leq n+|J_0|-1, \\ J_0 \subset J}} S_{J_0}^{SOC}(k) \; .$$

Each of these sets S_1, S_2, S_3 is closed. So also the union

$$S := S_1 \cup S_2 \cup S_3$$

is a closed semi-algebraic subset of $\mathcal{N} = \mathbb{R}^m \times \big(\mathbb{R}^n\big)^{m+1} \times \big(S^n\big)^{m+1}$ with

$$\text{codim } S \geq n + 1 \; .$$

This closed semi-algebraic set S allows a Whitney regular stratification (see, Theorem 5.9)

$$(S, \Sigma) \; , \; \Sigma = \{X_i, \; i \in I\} \; , \quad \text{with finitely many strata} X_i, i \in I \; ,$$

and we can define the set $\mathcal{P}_r$ of nice problems $P(t) = P(f,G)$ (cf., (6.37))

$$\mathcal{P}_r = \{(f,G) \in C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R})^{m+1} \mid \tilde{j}^2(f,G) \pitchfork X_i \; \forall X_i \in \Sigma\} \; .$$

According to the jet-transversality Theorem 6.2 we obtain

COROLLARY 6.4. *The set $\mathcal{P}_r$ is open and dense in $C^\infty(\mathbb{R}^n \times \mathbb{R}^p, \mathbb{R})^{m+1}$ wrt. the C^k_s-topology for all $k \geq 3$.*
The sets $\left(\tilde{j}^2(f,G)\right)^{-1}(X_i)$ are manifolds in $\mathbb{R}^n \times \mathbb{R}^p$ with codimension $c_d = \text{codim } X_i,\ i \in I$, satisfying $c_d \geq n+1$.

By definition, the set C_{gc} is a closed subset of $\mathbb{R}^n \times \mathbb{R}^p$, and as discussed above (cf., Figure 6.14), the set $C_r \subset C_{gc}$ of nondegenerate critical points is a manifold in $\mathbb{R}^n \times \mathbb{R}^p$ of dimension p (codimension n).
The set $C_s = C_{gc} \setminus C_r$ equals $\left(\tilde{j}^2(f,G)\right)^{-1}(S) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R}^p \mid \tilde{j}^2(f,G)(x,t) \in S\}$ and this set consists of manifolds $\left(\tilde{j}^2(f,G)\right)^{-1}(X_i)$, $X_i \in \Sigma$, of $\mathbb{R}^n \times \mathbb{R}^p$ of codimension $\geq n+1$. Recall that by definition the set C_s contains precisely the gc-points of $P(f,G)$ where (at least) one of the conditions LICQ, SC, or SOC fails.
We finish this subsection with some remarks on problems $P(t) = P(f,G)$ for (f,G) from the generic set $\mathcal{P}_r$.

- For $p = 1$ the mapping $\tilde{j}^2(f,G)$ avoids strata $X_i \in \Sigma$ with codim $X_i \geq n+2$ and possibly intersects the strata of codimension $n+1$ at a discrete set of gc-points (x,t) from C_s.
- We emphasize that until now, only for $p = 1$ a complete description of the generic structure of the set C_{gc} has been achieved. To obtain more details on generic properties for the case $p > 1$, we would need a more specific stratification $\Sigma = \{X_i,\ i \in I\}$ of the jet set $\mathcal{N}$ such that each stratum corresponds to a specific distribution of degeneracies a gc-point (x,t) may show in the generic situation. More precisely, we should define strata X_i which precisely correspond, e.g., to situations where LICQ and SC fails but not SOC, or where SC and SOC fails but not LICQ etc. Even for $p = 2$ this becomes very complex.

REMARK 6.2. Let us mention that the genericity results for parametric problems of the present chapter contains the genericity results for the corresponding non-parametric problems in Chapter 4 as a special case if we chose $p = 0$.

Unconstrained programs: Note that all arguments at the beginning of Subsection 6.3.1 remain true if we chose $p = 0$. For this choice the set $\mathcal{P}^1_r$ in Lemma 6.2 equals the set $\mathcal{P}_r$ in (4.6), and for $p = 0$ Lemma 6.2 coincides with Theorem 4.5.

Systems of nonlinear equations: Here the set $\mathcal{P}_r$ in Theorem 6.8 for $p = 0$ equals the set $\mathcal{P}_r$ in (4.10) and Theorem 6.8 coincides with Theorem 4.6.

Constrained programs: Also here, for $p = 0$ the statement of Corollary 6.4 yields the genericity result of Theorem 4.9 for non-parametric programs.

6.5.2. One-parametric constrained programs.

We are especially interested in parametric programs $P(t)$ (cf., (6.30)) with $t \in \mathbb{R}$, i.e., in the one-parametric case $p = 1$.
Here, the generic behavior of the sets C_{gc} of generalized critical points and the sets C_m corresponding to local minimizers is completely understood and analysed in the original papers of Jongen et al. [23, 22] and the books [21, 17]. We do not go into technical details

and only summarize the generic properties as presented in [17, Chapter 2], and offer some illustrative examples.

For so-called pathfollowing methods to trace numerically the curves of gc-points of one-parametric programs $P(t)$, we also refer to [17].

As we shall see, generically in the sets C_{gc} only 5 types of gc-points can appear, namely nondegenerate critical points (Type 1), and points of Types 2-5, where precisely one of the regularity conditions SC, SOC, LICQ fails. We roughly describe the generic behavior of C_{gc} (or C_m) near a gc-point $(\overline{x}, \overline{t})$ of $P(t) = P(f, G)$ for each type and finally present the genericity results.

Type 1: These are nondegenerate critical points $(\overline{x}, \overline{t})$ (see Definition 6.2) where LICQ, SC, and SOC are satisfied. We have seen above that near such a point $(\overline{x}, \overline{t})$ the set C_{gc} consists of a one-dimensional manifold given by a C^∞-curve $(x(t), t),\ t \approx \overline{t}, t \in \mathbb{R}$ (see Figure 6.15).

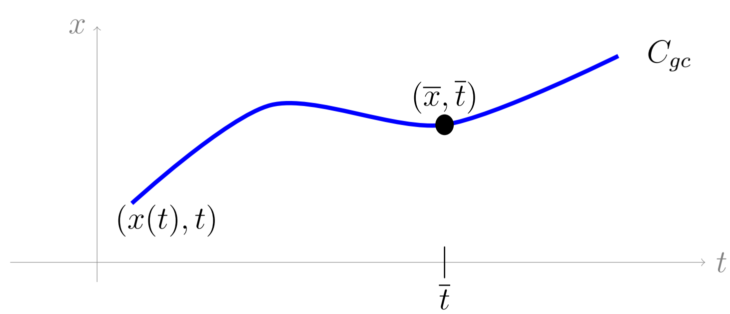

FIGURE 6.15. Sketch of C_{gc} near a point $(\overline{x}, \overline{t})$ of $P(t)$ of Type 1.

In case that for the point $(\overline{x}, \overline{t})$ of Type 1 the vector $\overline{x}$ is a nondegenerate local minimizer of $P(\overline{t})$, locally near $(\overline{x}, \overline{t})$, this curve $(x(t), t)$ consists of nondegenerate local minimizers $x(t)$ of $P(t)$.

Type 2: These are critical points $(\overline{x}, \overline{t})$, where SC fails but LICQ and SOC holds. More precisely we have

(2a) $J_{(\overline{x}, \overline{t})} \neq \emptyset$, and precisely one of the multipliers in (6.33) is equal to zero, say:

$$\overline{\mu}_p = 0, \qquad \text{for one } \ p \in J_{(\overline{x}, \overline{t})} \ .$$

(2b) and 3 other specific conditions are satisfied.
All these conditions depend explicitly on quantities which are completely determined by $(\overline{x}, \overline{t})$ and f, G (see [17, p. 43] for details).

Near such a point $(\overline{x}, \overline{t})$ of Type 2 there are two C^∞-curves $(x(t), t)$ and $(\tilde{x}(t), t)$, $t \approx \overline{t}$ (see Figure 6.16), where the first curve (in blue) consists of critical points with $J_{(x(t),t))} = J_{(\overline{x}, \overline{t})}$ and with a corresponding C^∞-multiplier vector $\mu(t)$. The component $\mu_p(t)$ changes from positive to negative (or from negative to positive) when passing $t = \overline{t}$ on this curve.

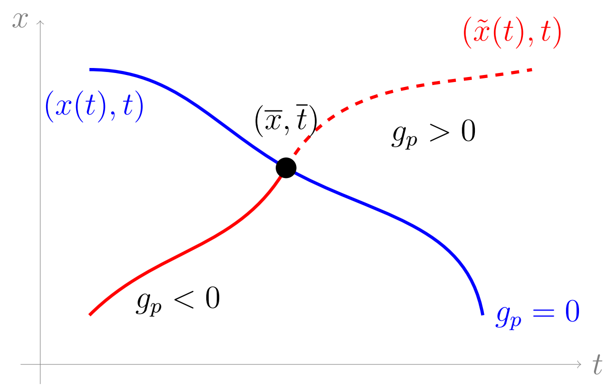

FIGURE 6.16. Sketch of C_{gc} near a point $(\overline{x}, \overline{t})$ of Type 2.

The other curve $(\tilde{x}(t), t)$ (in red) consists of gc-points of the program $\tilde{P}(t)$ obtained from $P(t)$ by skipping the constraint $g_p(x, t) \leq 0$. Half of this curve (red line segment) satisfies $g_p(\tilde{x}(t), t)) < 0,\ t < \overline{t}$ (or $t > \overline{t}$), and thus also belongs to C_{gc}. But the other red dashed part satisfies $g_p(\tilde{x}(t), t)) > 0,\ t > \overline{t}$ (or $t < \overline{t}$), and thus does not belong to C_{gc} (is not feasible for $P(t)$).
So generically near a critical point of type 2, the set C_{gc} looks as sketched in Figure 6.17.

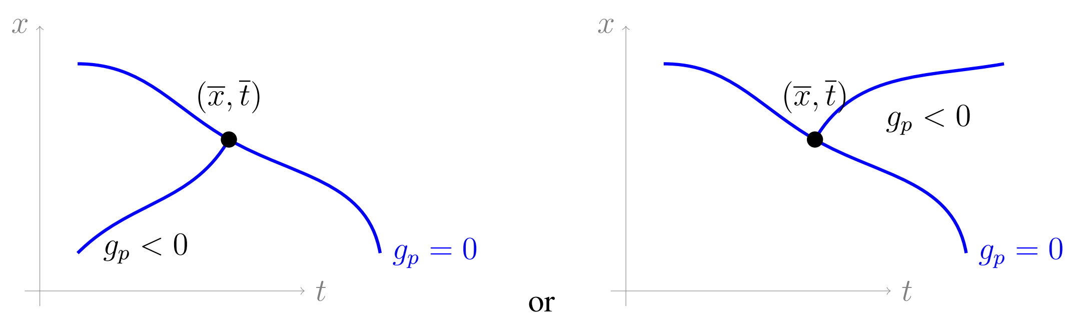

FIGURE 6.17. Set C_{gc} near a point $(\overline{x}, \overline{t})$ of Type 2.

Moreover, if on one branch of the set C_{gc} there are local minimizers, then near $(\overline{x}, \overline{t})$, the set C_m looks topologically as in Figure 6.18.

Type 3: At critical points $(\overline{x}, \overline{t})$ of this type the condition SOC fails but LICQ and SC holds. More precisely, the critical point equation $\nabla_x L(\overline{x}, \overline{t}, \overline{\mu}) = 0$ (cf., (6.33)) holds such that with $\overline{V} = [\overline{v}_1 \ldots \overline{v}_k]$, where $\overline{v}_j,\ j = 1, \ldots, k,\ k = n - |J_{(\overline{x},\overline{t})}|$, is a basis of the tangent space $T_{(\overline{x},\overline{t})}$ (cf., Definition 6.2), it follows:

(3a) $\overline{V}^T \nabla_x^2 L(\overline{x}, \overline{t}, \overline{\mu}) \overline{V}$ has exactly one zero eigenvalue,

(3b) and one further specific condition is satisfied.
This condition is completely determined by $(\overline{x}, \overline{t})$ and f, G (see [17, p. 44f.] for details).

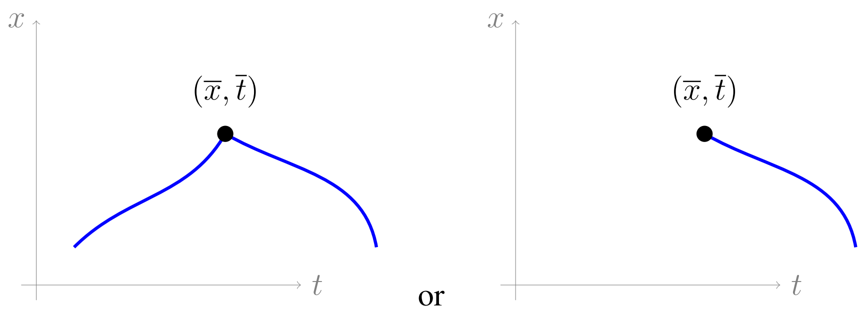

FIGURE 6.18. Set C_m near a point $(\overline{x}, \overline{t})$ of Type 2.

Near such a point $(\overline{x}, \overline{t})$ of Type 3 the set C_{gc} looks like sketched in Figure 6.19, i.e., $(\overline{x}, \overline{t})$ is a turning point. On these curves, precisely one eigenvalue of $V^T \nabla_x^2 L(x, t, \mu) V$ changes from positive to negative (or from negative to positive) when passing $(\overline{x}, \overline{t})$.

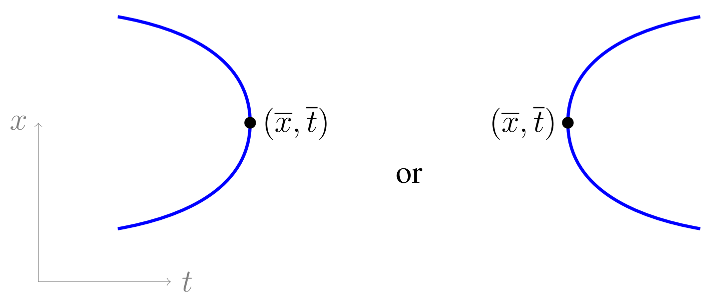

FIGURE 6.19. Sketch of C_{gc} near a point $(\overline{x}, \overline{t})$ of Type 3.

Consequently, since for nondegenerate local minimizers all eigenvalues of $V^T \nabla_x^2 L(x, t, \mu) V$ must be positive, near a Type 3 critical point $(\overline{x}, \overline{t})$, (in case one ”side” of C_{gc} corresponds to local minimizers) the set C_m must topologically look like in Figure 6.20. Note that the point $(\overline{x}, \overline{t})$ need not belong to C_m.

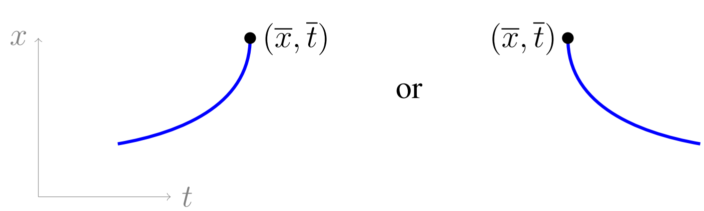

FIGURE 6.20. Sketch of C_m near a point $(\overline{x}, \overline{t})$ of Type 3.

Type 4: For these critical points $(\overline{x}, \overline{t})$ LICQ fails with $|J_{(\overline{x}, \overline{t})}| \leq n$. More precisely it holds,

(4a) $0 < |J_{(\overline{x}, \overline{t})}| \leq n$, dim span $\{\nabla_x g_j(\overline{x}, \overline{t}),\ j \in J_{(\overline{x}, \overline{t})}\} = |J_{(\overline{x}, \overline{t})}| - 1$,

(4b) and 3 further specific conditions are satified.
These conditions are completely determined by $(\overline{x}, \overline{t})$ and f, G (see [17, p. 46f.] for details).

Near such a point $(\overline{x}, \overline{t})$ of Type 4 the set C_{gc} looks like given in Figure 6.21, i.e., $(\overline{x}, \overline{t})$ is a turning point.

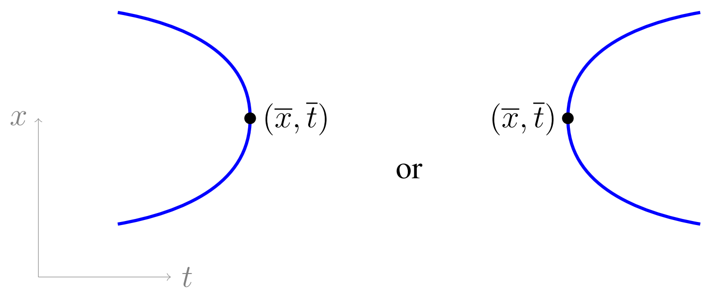

FIGURE 6.21. Sketch of C_{gc} near a point $(\overline{x}, \overline{t})$ of Type 4.

On these curves the signs of all multipliers $\mu_j,\ j \in J_{(\overline{x},\overline{t})}$, and the signs of all eigenvalues of $V^T \nabla_x^2 L(x, t, \mu) V$ change (V as in Type 3). So as for Type 3 (in case one "side" corresponds to local minimizers) the set C_m must topologically look like in Figure 6.22, where the point $(\overline{x}, \overline{t})$ need not belong to C_m.

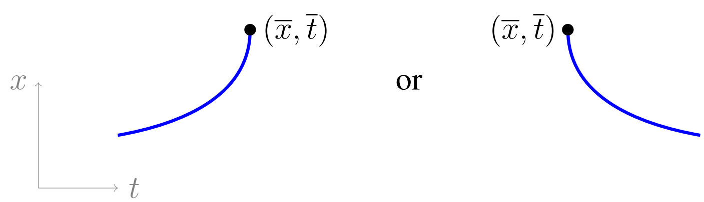

FIGURE 6.22. Sketch of C_m near a point $(\overline{x}, \overline{t})$ of Type 4.

Type 5: These are gc-points $(\overline{x}, \overline{t})$ where LICQ fails with $|J_{(\overline{x},\overline{t})}| = n+1$. More precisely we have,

(5a) Since $|J_{(\overline{x},\overline{t})}| = n+1$, the gradients $\nabla_x g_j(\overline{x}, \overline{t}),\ j \in J_{(\overline{x},\overline{t})}$, are linearly dependent,

(5b)

$$\nabla_{(x,t)} g_j(\overline{x}, \overline{t}),\ j \in J_{(\overline{x},\overline{t})}, \quad \text{are linearly independent,} \tag{6.44}$$

(5c) and 2 further specific conditions are satisfied.
These conditions are completely determined by $(\overline{x}, \overline{t})$ and f, G (see [17, p. 48f.] for details).

By using (6.44) and $|J_{(\overline{x},\overline{t})}| = n + 1$ one can show that for each $q \in J_{(\overline{x},\overline{t})}$ the point $(\overline{x}, \overline{t})$ of Type 5 is a nondegenerate critical point of the program $P_q(t)$ obtained from $P(t)$ by skipping the constraint $g_q(x,t) \leq 0$. So near $(\overline{x},\overline{t})$, for each $q \in J_{(\overline{x},\overline{t})}$ there exists a curve $(x_q(t),t),\ t \approx \overline{t}$, of nondegenerate critical points of $P_q(t)$ (see Figure 6.23). One "side"of this curve correspond to points satisfying $g_q(x,t) < 0$ (in blue) whereas on the other "side" (in red) the condition $g_q(x.t) > 0$ is fulfilled, so that these points of the curve are not feasible for $P(t)$. Thus only the blue part together with $(\overline{x},\overline{t})$ belong to C_{gc}.

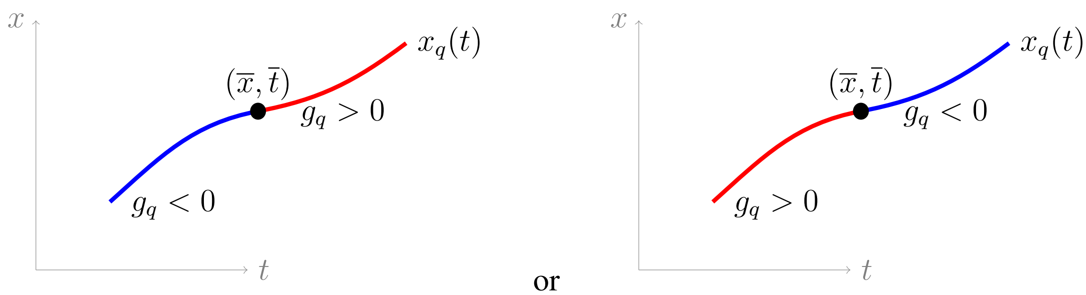

FIGURE 6.23. Sketch of the curve $(x_q(t),t)$ near a point $(\overline{x},\overline{t})$ of Type 5.

Any $q \in J_{(\overline{x},\overline{t})}$ gives rise to such a curve segment in C_{gc} (see Figure 6.24). The local picture of C_{gc} now depends on whether the Mangasarian Fromovitz Constraint Qualification is satisfied at $(\overline{x},\overline{t})$ or not.

DEFINITION 6.3. *The Mangasarian Fromovitz Constraint Qualification (MFCQ) is said to hold at $(\overline{x},\overline{t})$, $\overline{x} \in \mathcal{F}(\overline{t})$, $\overline{t} \in \mathbb{R}$, if there exists a vector $\xi \in \mathbb{R}^n$ satisfying,*

$$\nabla_x g_j(\overline{x},\overline{t})\xi < 0 \ , \qquad j \in J_{(\overline{x},\overline{t})} \ .$$

Note that MFCQ is weaker than LICQ.

If $(\overline{x},\overline{t})$ is a point of Type 5 satisfying MFCQ then near $(\overline{x},\overline{t})$ the set C_{gc}, resp, C_m looks like sketched in Figure 6.24, resp., Figure 6.25.

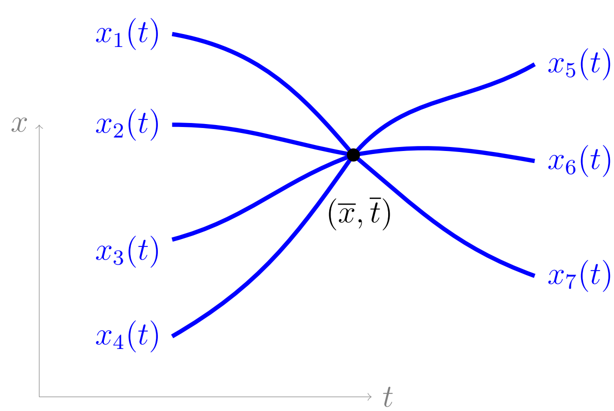

FIGURE 6.24. Sketch of C_{gc} near a point $(\overline{x},\overline{t})$ of Type 5 where MFCQ holds.

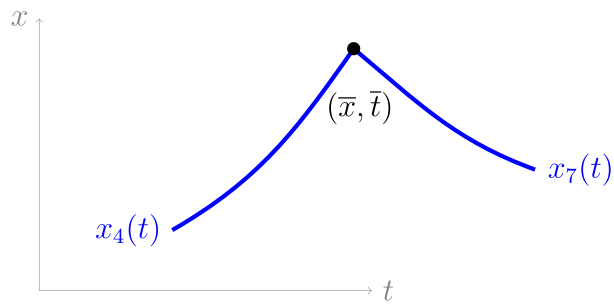

FIGURE 6.25. Sketch of C_m near a point $(\overline{x}, \overline{t})$ of Type 5 where MFCQ holds.

If MFCQ fails at $(\overline{x}, \overline{t})$ then this point is an "endpoint" of C_{gc} and C_m and these sets have the form as sketched in Figure 6.26, Figure 6.27. Concerning C_m, for this Type 5 the point $(\overline{x}, \overline{t})$ belongs to C_m.

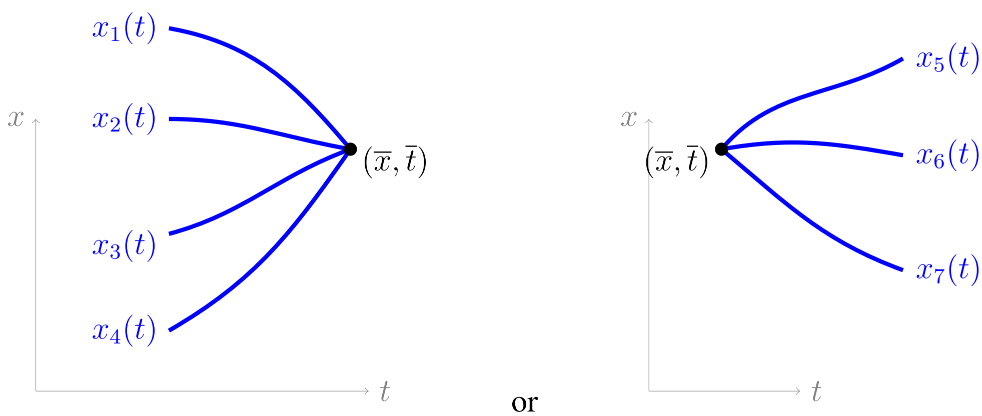

FIGURE 6.26. Sketch of C_{gc} near a point $(\overline{x}, \overline{t})$ of Type 5 where MFCQ fails.

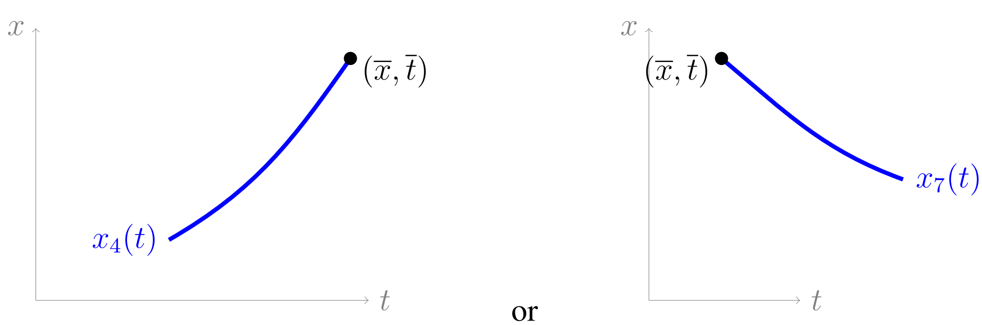

FIGURE 6.27. Sketch of C_m near a point $(\overline{x}, \overline{t})$ of Type 5 where MFCQ fails.

Now we summarize the genericity results for one-parametric problems $P(t) = P(f, G)$. Let us introduce the set

$\mathcal{P}_r = \{(f, G) \in [C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})]^{m+1} \mid$ each point in C_{gc} belongs to one of the Types 1-5.$\}$

We also define the sets

$$C^i_{gc} = \{(x,t) \in C_{gc} \mid (x,t) \text{ is of Type i}\}\ , \qquad i = 1,\dots,5\ .$$

THEOREM 6.10. [Genericity result of the 5 Types]
(a) The set $\mathcal{P}_r$ is open and dense in $[C^\infty(\mathbb{R}^n \times \mathbb{R}, \mathbb{R})]^{m+1}$ wrt. the C^k_s-topology for each $k \geq 3$.
(b) For $(f,G) \in \mathcal{P}_r$ the set C^1_{gc} is dense and (relatively) open in C_{gc}. Moreover, all sets $C^i_s,\ i = 2,\dots,5$, are discrete point sets.

Proof. Part (a) is Theorem 2.1 in [22]. Part (b) coincides with Theorem 2.2 in [22], and as argued in [22], this statement is a direct implication of the local structure of the set C_{gc} near a point of Type 1-5 as discussed above.

□

We finish the subsection with some additional observations on gc-points of the one-parametric program $P(t) = P(f,G)$.

- By Definition 6.1 the set C_{gc} of gc-points is always closed, and the (relatively) open set C^1_{gc} is a one-dimensional manifold in $\mathbb{R}^n \times \mathbb{R}$. So, for $(f,G) \in \mathcal{P}_r$, by Theorem 6.10(b), the set $C_{gc} \setminus C^1_{gc}$ consists of a discrete point set:
$$C_{gc} \setminus C^1_{gc} = \cup_{i=2,\dots,5}\ C^i_{gc}\ .$$
- Note that C_{gc} is closed but the set C_m need not be closed. Recall, e.g., the program (unconstrained, $J = \emptyset$)
$$P(t) \qquad \min_{x\in\mathbb{R}}\ f(x,t) := \frac{1}{3}x^3 - tx, \qquad t \in \mathbb{R}\ ,$$
with $C_{gc} = \{(x,t) \mid x^2 = t,\ t \geq 0\}$ (see Figure 6.28).

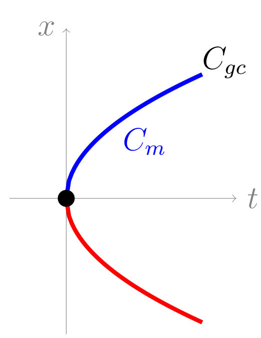

FIGURE 6.28. Critical point set C_{gc} for $f(x,t) := \frac{1}{3}x^3 - tx$ with turning point $(\overline{x},\overline{t}) = (0,0)$.

The upper part (in blue) consists of points (x,t) with minimizers x of $P(t)$ ($\nabla^2_x f(x,t) = 2x > 0$) and the points on the lower part (in red) correspond to maximizers ($\nabla_x^2 f(x,t) = 2x < 0$). But the turning point $(\overline{x}, \overline{t}) = (0,0)$ is only a critical point (for $\min \frac{1}{3}x^3$) and $\overline{x}$ is neither a minimizer nor a maximizer of $P(0)$.

To illustrate the mecanism of the generic structure of the sets C_{gc} and C_m of a one-parametric program $P(t) = P(f,G)$ near a gc-point $(\overline{x}, \overline{t})$ of Type 2,3,4 and 5, we present some typical examples for $x \in \mathbb{R}^2,\ t \in \mathbb{R}$. These examples will not be given explicitly but visualized by sketches of the changes in the feasible sets (in blue) and the level sets of f (in red, dashed) for t near $\overline{t}$. We always indicate the situation for a point $t < \overline{t}$, for $t = \overline{t}$, and a value $t > \overline{t}$. A point x in $\mathbb{R}^2$, numbered, e.g., by 1, 2, or 3, always belongs to a gc-point (x,t) of $P(t)$ on a branch of the set C_{gc} (or C_m) with the same number (see, e.g., Figure 6.29 and Figure 6.30 (left)). The direction $-\nabla f$ indicates the direction of the negative gradient of f on a point of the level line of f.

<u>Examples for Type 2:</u>

A first example is given by the following sketches in $\mathbb{R}^2$ of the feasible sets, with 2 constraints $g_1, g_2 \leq 0$, and a level set of f. As the intersection of two active constraints

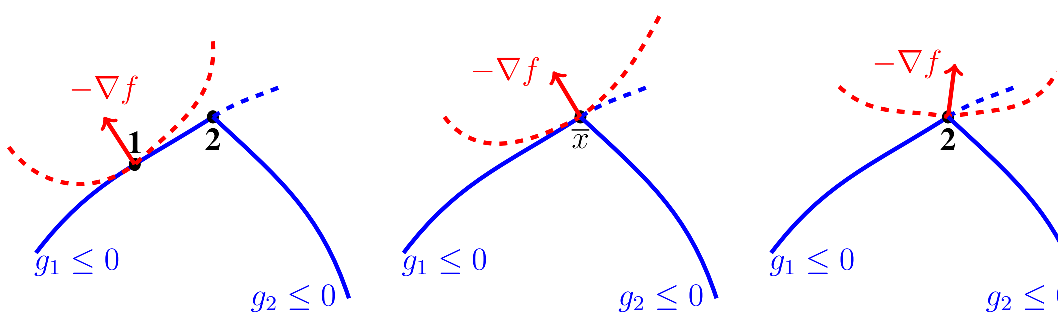

FIGURE 6.29. Picture for: $t < \overline{t}$, $t = \overline{t}$, $t > \overline{t}$.

($g_1 = g_2 = 0$) in $\mathbb{R}^2$ the point 2 represents a gc-point. So, these pictures correspond to a set C_{gc} and C_m in $\mathbb{R}^2 \times \mathbb{R}$ of the form given in Figure 6.30.

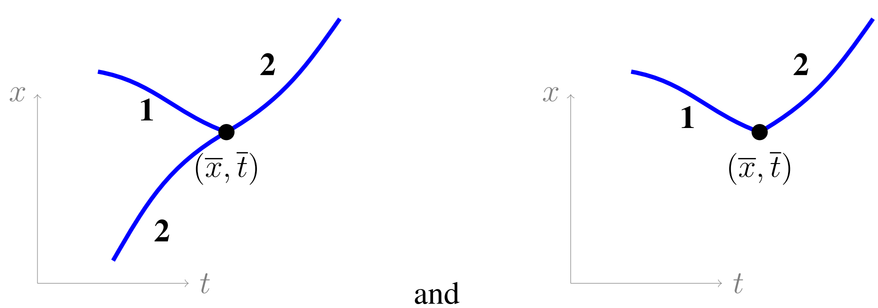

FIGURE 6.30. Sketch of C_{gc} (left), C_m (right) near the gc-point $(\overline{x}, \overline{t})$ of Type 2.

In the second example of Type 2, the objective function is (up to small C^∞-perturbations) of the form $f(x_1, x_2) = x_1 x_2$ with level lines as in Figure 6.31.

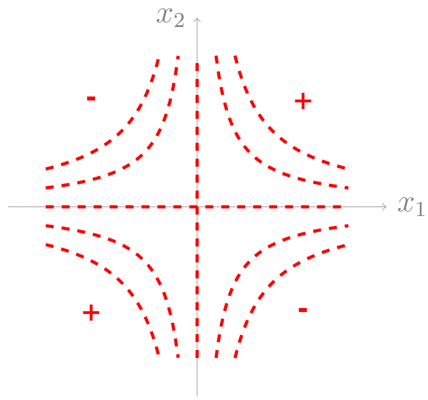

FIGURE 6.31. Some level lines for $f(x) = x_1 x_2$.

In this example we are given one constraint $g \leq 0$ and the program is sketched in Figure 6.32. Here, at point 2, the condition $\nabla_x f(x,t) = 0$ is always satisfied, so that this

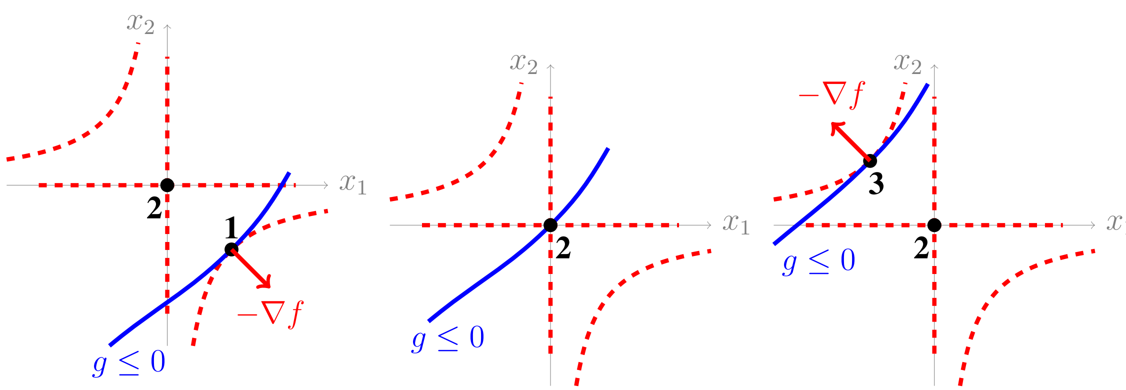

FIGURE 6.32. Picture for: $t < \bar{t}$, $t = \bar{t}$, $t > \bar{t}$.

point (if feasible) belongs to a gc-point (neither a minimizer nor a maximizer). With the orientation of $-\nabla f$, point 1 is a critical point (neither a minimizer nor a maximizer) and point 3 is a local minimizer. So, for this problem the set C_{gc} and C_m has the form given in Figure 6.33.

Example for Type 3:

A typical example has been presented before in Section 6.3 by the unconstrained minimization problem ($J = \emptyset$) $\min \; f(x,t) = \frac{1}{3}x^3 - tx$, with $(x,t) \in \mathbb{R} \times \mathbb{R}$ (see the sketches of the set of critical points $C(f) = C_{gc}$ in Figure 6.6).

Example for Type 4:

In the first example MFCQ is satisfied at the gc-point $(\bar{x}, \bar{t})$, in the second MFCQ fails. In both examples we are given two constraints $g_1 \leq 0$ and $g_2 \leq 0$.

Consider the sketches in Figure 6.34.

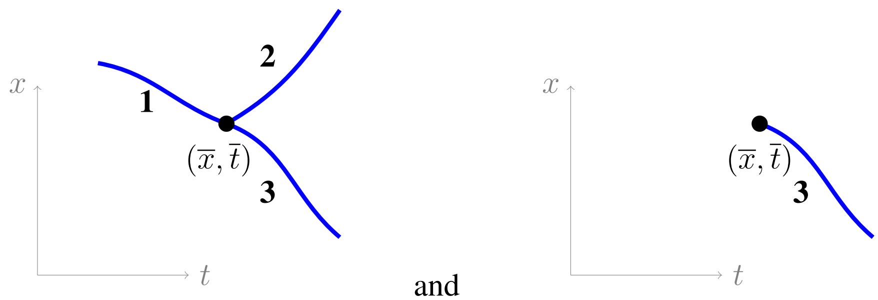

FIGURE 6.33. Sketch of C_{gc} (left), C_m (right) near the gc-point $(\overline{x}, \overline{t})$ of Type 2.

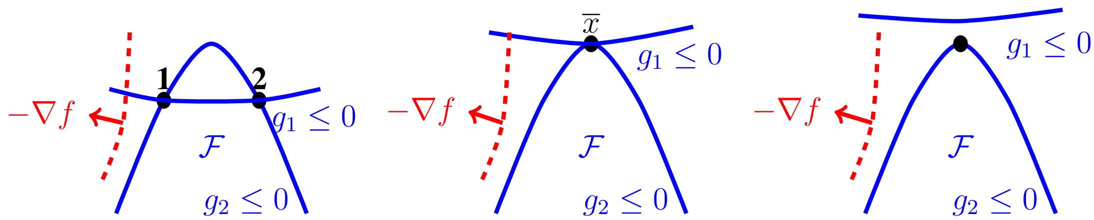

FIGURE 6.34. Picture for: $t < \overline{t}$, $t = \overline{t}$, $t > \overline{t}$.

For $t < \overline{t}$ the points 1 and 2 are intersection points of two linear independent constraints ($g_1 = g_2 = 0$), and they therefore correspond to critical points. Also for $t = \overline{t}$, the point $(\overline{x}, \overline{t})$ is a generalized critical point. With the indicated orientation of $-\nabla f$ the points 1,2 (and $\overline{x}$) do neither belong to minimizers nor to maximizers. For $\overline{t} < t$ there is no gc-point (x, t) near $(\overline{x}, \overline{t})$. So in a neighborhood of $(\overline{x}, \overline{t})$ the set C_{gc} looks like sketched in Figure 6.35, and C_m is empty.

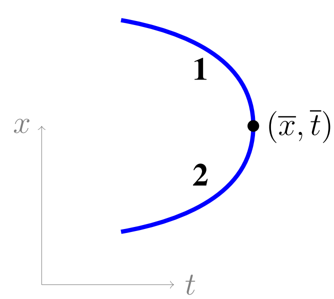

FIGURE 6.35. Sketch of C_{gc} near the gc-point $(\overline{x}, \overline{t})$ of Type 4.

In the second example program $P(t)$ the condition MFCQ fails at $(\overline{x}, \overline{t})$ (see Figure 6.36). For $t < \overline{t}$, with the indicated orientation of $-\nabla f$, the point 1 corresponds to a local minimizer and point 2 represents a local maximizer. For $t = \overline{t}$, the point $\overline{x}$ is an isolated feasible point and thus $\overline{x}$ is a local minimizer (and maximizer) of $P(\overline{t})$. For $t > \overline{t}$, near $\overline{x}$ the feasible set of $P(t)$ is empty. So for this example the picture of C_{gc}, C_m is as in Figure 6.37.

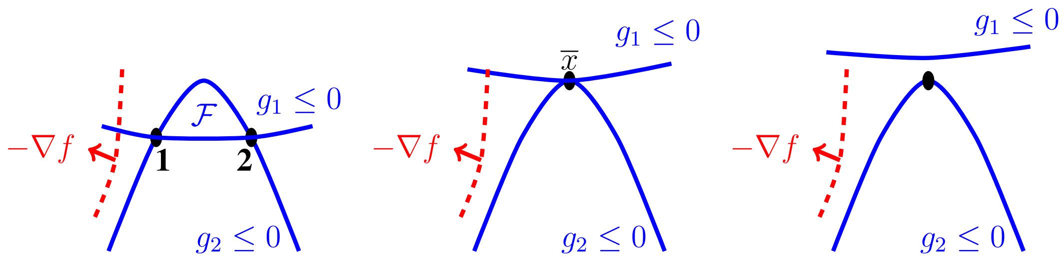

FIGURE 6.36. Picture for: $t < \bar{t}$, $t = \bar{t}$, $t > \bar{t}$.

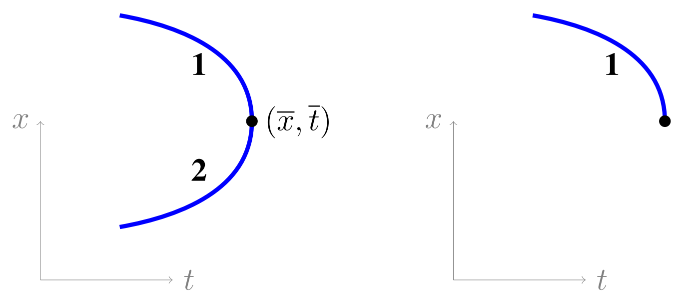

FIGURE 6.37. Sketch of C_{gc} (left) and C_m (right) near the gc-point $(\overline{x}, \overline{t})$ of Type 4.

Example for Type 5:

We again present sketches of two example programs. In the first example MFCQ is satisfied at the gc-point $(\overline{x}, \overline{t})$, in the second not. In both examples we are given three constraints $g_1 \leq 0, g_2 \leq 0$, and $g_3 \leq 0$.

Consider the pictures in Figure 6.38. With this orientation of $-\nabla f$ the point 1 belongs to

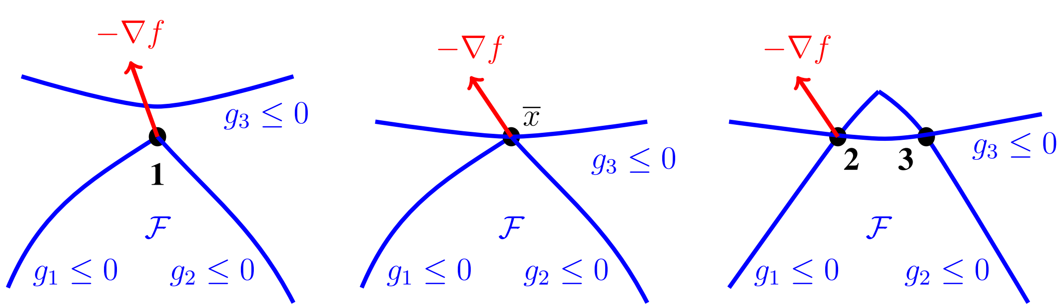

FIGURE 6.38. Picture for: $t < \bar{t}$, $t = \bar{t}$, $t > \bar{t}$.

a local minimizer for $t < \bar{t}$ and $t = \bar{t}$. For $\bar{t} < t$ point 2 belongs to a local minimizer and point 3 to a critical point. So, in a neighborhood of $(\overline{x}, \overline{t})$ the sets C_{gc} and C_m look like presented in Figure 6.39.

Now look at the second example program $P(t)$ where MFCQ fails at $(\overline{x}, \overline{t})$, see Figure 6.40. For $t < \bar{t}$, the feasible set and thus C_{gc} is (locally) empty. For $t = \bar{t}$, the

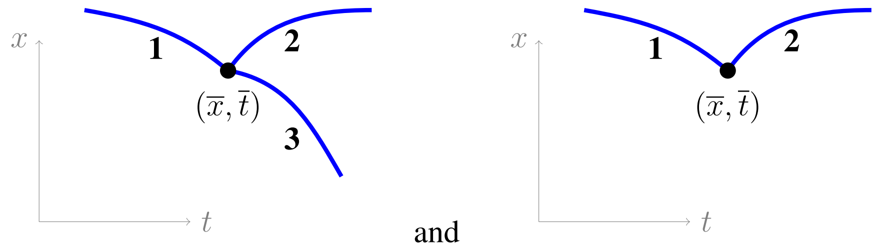

FIGURE 6.39. Sketch of C_{gc} (left), C_m (right) near the gc-point $(\overline{x}, \overline{t})$ of Type 5 where MFCQ holds.

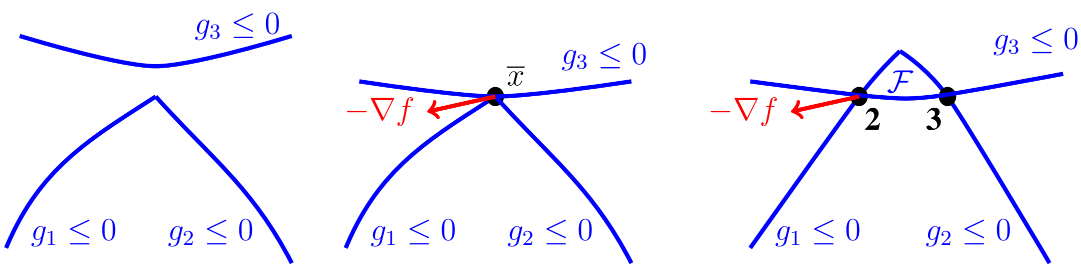

FIGURE 6.40. Picture for: $t < \overline{t}$, $t = \overline{t}$, $t > \overline{t}$.

isolated feasible point $(\overline{x}, \overline{t})$ is a local minimizer and maximizer, and with the indicated orientation of $-\nabla f$, for $t > \overline{t}$, the point 2 corresponds to a local minimizer and point 3 represents a local maximizer. So, the picture of C_{gc}, C_m is as in Figure 6.41.

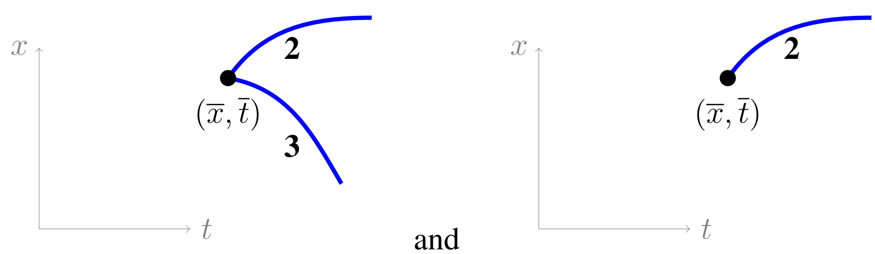

FIGURE 6.41. Sketch of C_{gc} (left), C_m (right) near the gc-point $(\overline{x}, \overline{t})$ of Type 5 where MFCQ fails.

6.6. One-parametric quadratic and one-parametric linear programs

In this section we examine two special classes of parametric constrained programs, namely one-parametric quadratic and one-parametric linear programs. The aim is to discuss the differences between the generic behaviour of one-parametric general constrained programs $P(t)$ (cf., (6.30)) analysed in Section 6.5.2 (results of the 5 Types) and the generic structure for one-parametric quadratic or linear programs. The results in particular show again, that genericity results obtained for a class of problems need no more be valid for a subclass of problems with a special substructure. Unfortunately, any special subclass needs a new specific genericity analysis.

6.6.1. Genericity results for one-parametric quadratic programs. In this subsection we consider one-parametric problems of the type,

$$(6.45)\qquad Q(t): \quad \min\ f(x,t) := \frac{1}{2}x^T C(t)x + c(t)^T x \quad \text{s.t.} \quad x \in \mathcal{F}(t) = \{x \in \mathbb{R}^n \mid B(t)x \le b(t)\}$$

depending on the parameter $t \in \mathbb{R}$, where $C \in C^\infty(\mathbb{R}, S^n)$, $c \in C^\infty(\mathbb{R}, \mathbb{R}^n)$, $B \in C^\infty(\mathbb{R}, M^{m,n})$, $b \in C^\infty(\mathbb{R}, \mathbb{R}^m)$. So, $Q(t)$ has a quadratic objective and m linear constraints

$$g_j(x,t) := B_{j\cdot}(t)x - b_j(t) \le 0\ , \quad j \in J := \{1,\ldots,m\}\ ,$$

and is given by the problem data

$$(6.46)\qquad (C,c,B,b) \in C^\infty(\mathbb{R}, S^n \times \mathbb{R}^n \times \mathbb{R}^{mn} \times \mathbb{R}^m) \equiv C^\infty(\mathbb{R}, \mathbb{R}^N)\ ,$$

where $N = \frac{1}{2}(n+3)n + (n+1)m$. Often we write $Q(C,c,B,b)$ instead of $Q(t)$. Recall that a point $(\overline{x},\overline{t})$, $\overline{x} \in \mathcal{F}(\overline{t})$, is a gc-point of $Q(t)$ if the gradients

$$\nabla_x^T f(\overline{x},\overline{t}) = C(\overline{t})x + c(\overline{t})\ , \quad \nabla_x^T g_j(\overline{x},\overline{t}) = B_{j\cdot}(\overline{t})^T,\ j \in J_{(\overline{x},\overline{t})},\ \text{are linearly dependent}\ .$$

LICQ holds at $(\overline{x},\overline{t})$ if the vectors $B_{j\cdot}(\overline{t})^T$, $j \in J_{(\overline{x},\overline{t})}$, are linearly independent. Then at a gc-point $(\overline{x},\overline{t})$ satisfying LICQ the critical point relation holds:

$$(6.47)\qquad C(\overline{t})\overline{x} + c(\overline{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \overline{\mu}_j B_{j\cdot}(\overline{t})^T = 0\ ,\ \text{with unique } \overline{\mu}_j\ .$$

As usual we say that at a critical point $(\overline{x},\overline{t})$

$$(6.48)\qquad \text{(SC) holds, if in (6.47):} \qquad \overline{\mu}_j \neq 0 \quad \text{for all } j \in J_{(\overline{x},\overline{t})}\ ,$$

and with $T_{(\overline{x},\overline{t})} := \{d \in \mathbb{R}^n \mid B_{j\cdot}(\overline{t})d = 0,\ j \in J_{(\overline{x},\overline{t})}\}$

$$(6.49)\qquad \text{(SOC) holds, if:} \quad d^T C(\overline{t})d \neq 0 \quad \text{for all } d \in T_{(\overline{x},\overline{t})} \setminus \{0\}\ .$$

As before, we are interested in the generic structure of the set C_{gc} of gc-points of $Q(t)$. This structure has been completely described in [27]. Here we will only present the main results mostly without proofs. It appears that generically for $Q(t)$ only gc-points of Type 1,2,5 occur.

Type 1: A gc-point $(\overline{x},\overline{t})$, where LICQ, SC, and SOC holds.
The set of all such critical points is denoted by C_{gc}^1.

Type 2: A gc-point $(\overline{x},\overline{t})$, where LICQ holds but SC fails with precisely one multiplier equal to zero: $\overline{\mu}_p = 0$, for one $p \in J_{(\overline{x},\overline{t})}$.
Moreover two other conditions hold (see [27, Definition 2.2]).

Type 5: A gc-point $(\overline{x},\overline{t})$, where LICQ fails with $|J_{(\overline{x},\overline{t})}| = n+1$.
Moreover 3 other conditions hold (see [27, Definition 2.3]).

These types are the same as for the general constrained programs $P(t)$ in Subsection 6.5.2. Remind that as for general programs in Subsection 6.5.2, the set C_{gc}^1 of Type 1 gc-points forms a one dimensional C^∞-manifold (locally) parameterized by a C^∞-curve $(x(t),t)$.

However, different from the general programs $P(t)$, here, for quadratic problems $Q(t)$, the following may happen: There is a point $\hat{t} \in \mathbb{R}$ and a branch of C^1_{gc} consisting of a curve locally given by $(x(t), t)$ which satisfies

$$\lim_{t \uparrow \hat{t}} \|x(t)\| = \infty \quad \text{or} \quad \lim_{t \downarrow \hat{t}} \|x(t)\| = \infty \; .$$

As will be shown below, generically for $Q(t)$ we may see a behavior as in Figure 6.42 at such points $\hat{t}$, which will be called

points $\hat{t}$ of "Type 3 at infinity" and of "Type 4 at infinity" .

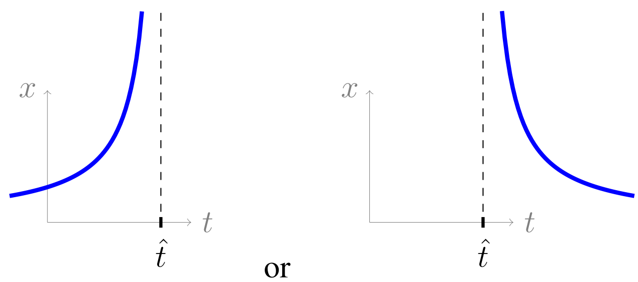

FIGURE 6.42. Point $\hat{t}$ of Type 3 and of Type 4 at infinity.

First we give the genericity result for the Types 1,2,5.

THEOREM 6.11. [Genericity result of the 3 Types]
(a) There is a subset $\mathcal{P}_r$ of $C^\infty(\mathbb{R}, \mathbb{R}^N)$ (see (6.46)) which is open and dense in $C^\infty(\mathbb{R}, \mathbb{R}^N)$ wrt. the C^k_s-topology for any $k \geq 1$, such that the following is valid: For all $(C, c, B, b) \in \mathcal{P}_r$ any gc-point of $P(C, c, B, b)$ belongs to one of the Types 1,2 or 5.
(b) For $(C, c, B, b) \in \mathcal{P}_r$ the sets of gc-points of Type 2 and Type 5 are discrete point sets which are situated in the toplological closure of the manifold C^1_{gc}.

Proof. A detailed proof of this theorem is to be found in [27, Th. 3.1, Cor. 3.1]. Here we only comment on the main argument used.
Given the problem data (C, c, B, b) and any active index set $J_0 \subset J$ we define the matrix families
(6.50)

$$\mathcal{Q}(t) = \begin{pmatrix} C(t) & B(t)^T & c(t) \\ B(t) & 0 & -b(t) \end{pmatrix} \quad , \quad \mathcal{Q}_{J_0}(t) = \begin{pmatrix} C(t) & B_{J_0}(t)^T & c(t) \\ B_{J_0}(t) & 0 & -b_{J_0}(t) \end{pmatrix}$$

where as usual $B_{J_0} = (B_{j\cdot})_{j \in J_0}, \; b_{J_0} = (b_j)_{j \in J_0}$.
The genericity result is based on a Whitney regular stratification $(M_{n,q}, \Sigma)$ of the set $M_{n,q}$ of matrices M of the form (as $\mathcal{Q}_{J_0}(t)$ in (6.50), with fixed n, q, $q = |J_0|$):

$$M = \begin{pmatrix} C & D^T & c \\ D & 0 & d \end{pmatrix} \; , \quad C \in S^n, \; D \in M^{q,n}, \; c \in \mathbb{R}^n, \; d \in \mathbb{R}^q \; .$$

The strata set

$$\Sigma = \{V_i,\ i \in I\}\ , \qquad I \text{ finite}\ , \tag{6.51}$$

is tailored to the structure of gc-points occuring for $Q(t)$. This stratification allows the application of the transversality Theorem 6.2, which implies that for any $J_0 \subset J$, $|J_0| = q$, the set of families

$$\left\{ \mathcal{Q}_{J_0} \in C^\infty(\mathbb{R}, \mathbb{R}^{\frac{1}{2}(n+3)n+(n+1)q} \mid \mathcal{Q}_{J_0} \pitchfork V_i\ \forall i \in I \right\}$$

is C^k_s open and dense in $C^\infty(\mathbb{R}, \mathbb{R}^{\frac{1}{2}(n+3)n+(n+1)q})$ for $k \geq 1$.

□

We now describe the phenomenon leading to points $\hat{t}$ of Type 3 and of Type 4 at infinity. To do so, we consider a critical point $(\overline{x}, \overline{t})$ on a connected component of the one-dimensional manifold C^1_{gc}. Let this (maximal) component (branch) be parameterized by the C^∞-function $(x(t), t)$, $t \in I_{(\overline{x},\overline{t})}$, with a maximal open interval $I_{(\overline{x},\overline{t})}$ containing $\overline{t}$. We define

$$\hat{t} := \sup\{t \in I_{(\overline{x},\overline{t})}\}\ . \tag{6.52}$$

For $\hat{t} = \infty$ the picture for the right hand side of the branch is clear. For $\hat{t} < \infty$ the following may happen (cf., [27, Theorem 3.2]).

THEOREM 6.12. [Genericity result for points $\hat{t}$ of Type 3 and of Type 4 at infinity]
There is a subset $\mathcal{P}_r$ of $C^\infty(\mathbb{R}, \mathbb{R}^N)$ (see (6.46)*) satisfying the conditions in Theorem 6.11 as well as the following: For $(C, c, B, b) \in \mathcal{P}_r$ (cf., Theorem 6.11) let $\hat{t}$ be defined in* (6.52) *and satisfy $\hat{t} < \infty$. Then precisely one of the following alternatives (a) or (b) holds for $Q(C, c, B, b)$.*

(a) We have

$$\lim_{t \uparrow \hat{t}} x(t) = \hat{x} \quad \textit{for some } \hat{x} \in \mathbb{R}^n\ ,$$

and then $(\hat{x}, \hat{t})$ is a gc-point of Type 2 with $J_{(\overline{x},\overline{t})} = J_{(\hat{x},\hat{t})}$, $|J_{(\overline{x},\overline{t})}| \leq n$, or a gc-point of Type 5 with $J_{(\hat{x},\hat{t})} = J_{(\overline{x},\overline{t})} \cup \{j_0\}$ for some $j_0 \in J$, $|J_{(\hat{x},\hat{t})}| = n + 1$.

(b) We have

$$\lim_{t \uparrow \hat{t}} \|x(t)\| = \infty\ ,$$

and then precisely one of the following cases (i) or (ii) occurs for $\hat{t}$.

(i) *$\hat{t}$ is a point of Type 3 at infinity with $|J_{(\overline{x},\overline{t})}| < n$:*

Here we have with $J_0 = J_{(\overline{x},\overline{t})}$ (cf., (6.50)*)*

$$\text{rank} \begin{pmatrix} C(\hat{t}) & B_{J_0}(\hat{t})^T & c(\hat{t}) \\ B_{J_0}(\hat{t}) & 0 & -b_{J_0}(\hat{t}) \end{pmatrix} = \text{rank} \begin{pmatrix} C(\hat{t}) & B_{J_0}(\hat{t})^T \\ B_{J_0}(\hat{t}) & 0 \end{pmatrix} + 1 = n + |J_0|\ ,$$

and rank $B_{J_0}(\hat{t}) = |J_0|$.

(ii) *$\hat{t}$ is a point of Type 4 at infinity with $|J_{(\overline{x},\overline{t})}| = n$:*
Here we have with $J_0 = J_{(\overline{x},\overline{t})}$ (cf., (6.50))

$$\text{rank}\begin{pmatrix} C(\hat{t}) & B_{J_0}(\hat{t})^T & c(\hat{t}) \\ B_{J_0}(\hat{t}) & 0 & -b_{J_0}(\hat{t}) \end{pmatrix} = \text{rank}\begin{pmatrix} C(\hat{t}) & B_{J_0}(\hat{t})^T \\ B_{J_0}(\hat{t}) & 0 \end{pmatrix} + 1 = 2n\ ,$$

and rank $\left(B_{J_0}(\hat{t}), b_{J_0}(\hat{t})\right) = n$, rank $B_{J_0}(\hat{t}) = n-1$.

In both cases (i), (ii) we have

$$\lim_{t\uparrow\hat{t}} \frac{x(t)}{\|x(t)\|} = \tilde{x} \qquad \textit{with}\ \ \|\tilde{x}\| = 1\ .$$

More precisely, in case
(i) $\tilde{x}$ is given as $\tilde{x} = \frac{\hat{x}}{\|\hat{x}\|}$ where $(\hat{x},\hat{\mu})$ is the unique (up to a nonzero constant) solution $(\hat{x},\hat{\mu}) \neq (0,0)$ of

$$\begin{pmatrix} C(\hat{t}) & B_{J_0}(\hat{t})^T \\ B_{J_0}(\hat{t}) & 0 \end{pmatrix}\begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\ ,$$

and in case
(ii) $\tilde{x}$ is the unique (up to a sign) solution $\tilde{x}$ of $B_{J_0}(\hat{t})x = 0$, with $\|\tilde{x}\| = 1$.

The set of such points $\hat{t} \in \mathbb{R}$ in (a) and (b) is a discrete set.

We emphasize that a corresponding result is valid if we define

$$\hat{t} := \inf\{t \in I_{(\overline{x},\overline{t})}\} \quad \text{and consider} \quad \lim_{t\downarrow\hat{t}} x(t)\ .$$

For the points $\hat{t}$ of Type 3 and Type 4 at infinity we even can say more (cf., [27, Theorem 3.3]).

THEOREM 6.13. [Two-sided points $\hat{t}$ at infinity]
Let $(C,c,B,b) \in \mathcal{P}_r$ (cf., Theorem 6.12) and let $\hat{t}$ be a point of Type 3 or of Type 4 at infinity as in Theorem 6.12(b), so that the corresponding branch of C^1_{gc} left to $\hat{t}$ is given by the C^∞-curve $(x(t),t)$, $t < \hat{t}$, with active index set $J_{(x(t),t)} = J_{(\overline{x},\overline{t})}$ and satisfies $\lim_{t\uparrow\hat{t}} x(t)/\|x(t)\| = \tilde{x}$. Then
in Case $J_{(\overline{x},\overline{t})} = J$: there exists a branch of points in C^1_{gc} right to $\hat{t}$, parameterized by a C^∞-function $(x(t),t)$, $\hat{t} < t < \hat{t} + \varepsilon$ for some $\varepsilon > 0$, such that

$$\lim_{t\downarrow\hat{t}} \frac{x(t)}{\|x(t)\|} = -\tilde{x}\ .$$

Also for these Type 1 critical points $(x(t),t)$, $t > \hat{t}$, the active index set is the same, $J_{(x(t),t))} = J_{(\overline{x},\overline{t})}$, and for t $\downarrow \hat{t}$ the same Type occurs as for t $\uparrow \hat{t}$.
We call such points $\hat{t}$, two-sided points of Type 3 or of Type 4 at infinity.

In the reminder of the subsection we present examples to illustrate the phenomenon of points of Type 3 and Type 4 at infinity. We further explain the difference between the ordinary gc-point of Type 4 for general programs $P(t)$ and the Type 4 points at infinity for $Q(t)$.

Examples of points $\hat{t}$ of Type 3 at infinity:

Recall the simplest example of a Type 3 gc-point in Section 6.5, provided by the (non-quadratic) program (with $J = \emptyset$):

$$P(t): \qquad \min_{x\in\mathbb{R}} \frac{1}{3}x^3 - tx , \qquad t \in \mathbb{R},$$

with the critical point set C_{gc} near the critical point $(\overline{x}, \overline{t}) = (0,0)$ of Type 3 as sketched in Figure 6.6.

We now consider a one-parametric quadratic program with $J = \emptyset$ (see also [27, p.229]),

$$Q(t): \qquad \min_{x\in\mathbb{R}} \frac{1}{2}(1-t)x^2 + tx ,$$

with critical point equation $\nabla_x f(x,t) = (1-t)x + t = 0$. The set C_{gc} is thus given by

$$C_{gc} = \{(x(t), t) \mid x(t) = \frac{t}{t-1},\ t \neq 1\} ,$$

with a two-sided point $\hat{t} = 1$ of Type 3 at infinity (see Figure 6.43).

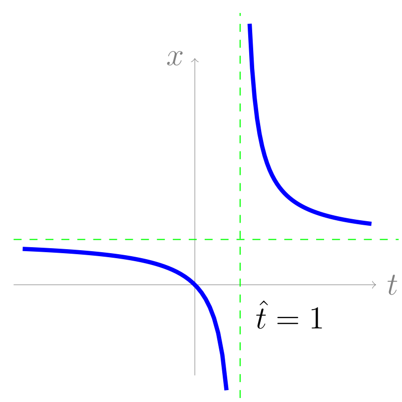

FIGURE 6.43. Two-sided point $\hat{t} = 1$ of Type 3 at infinity.

To see why (in the generic case) a two-sided point $\hat{t}$ of Type 3 at infinity can only occur if $J_{(\overline{x},\overline{t})} = J$, we now look at the problem obtained from $Q(t)$ by adding a constraint:

$$Q_1(t): \qquad \min_{x\in\mathbb{R}} \frac{1}{2}(1-t)x^2 + tx \quad \text{s.t.} \quad g(x,t) := x - t - 2 \leq 0 .$$

Here the constraint $x \leq t+2$ cuts off a part of the critical point set in Figure 6.43 given by $(x(t), t) = (\frac{t}{t-1}, t)$, $t \neq 1$. The activity condition $x = \frac{t}{t-1} = t+2$ or $t^2 = 2$ leads to two gc-points of Type 2 at $(x_1, t_1) = \big(2-\sqrt{2}, -\sqrt{2}\big)$ and $(x_2, t_2) = \big(2+\sqrt{2}, \sqrt{2}\big)$. Indeed, the KKT conditions $\nabla_x f(x,t) + \mu\nabla_x g(x,t) = 0$, $g(x,t) = 0$ read

$$(1-t)x + t + \mu = 0 , \quad x = t+2 \quad \text{or} \quad x = t+2 , \quad \mu = t^2 - 2 .$$

So the multiplier μ vanishes at $t = \pm\sqrt{2}$ and the points

$$(\tilde{x}(t), t) = (t + 2, t) , \qquad t \in \mathbb{R} ,$$

also belong to C_{gc}. The gc-points of $Q_1(t)$ are sketched in Figure 6.44. This set C_{gc} only shows a one-sided point $\hat{t}$ of Type 3 at infinity.

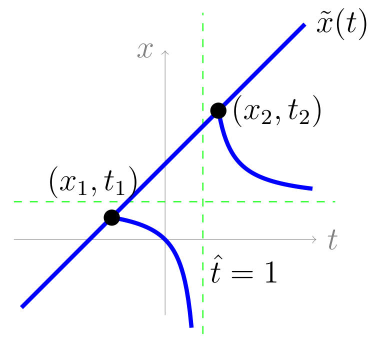

FIGURE 6.44. Set C_{gc} with a (one-sided) point $\hat{t} = 1$ of Type 3 at infinity and 2 critical points of Type 2 at (x_1, t_1), (x_2, t_2).

Examples of points $\hat{t}$ of Type 4 at infinity:

We end up the subsection with some explanation on points of Type 4 at infinity for the quadratic programs $Q(t)$. Consider first the example of a (general) program $P(t)$ with the (nonlinear) constraints for $x \in \mathbb{R}^2$,

$$g_1(x, t) = x_1^2 - x_2 \leq 0 , \quad g_2(x, t) = x_2 - t \leq 0 .$$

The feasible set $\mathcal{F}(t)$ (see Figure 6.45) is empty for $t < 0$, and consists of the isolated feasible point $\overline{x} = 0$ for $\overline{t} = 0$.

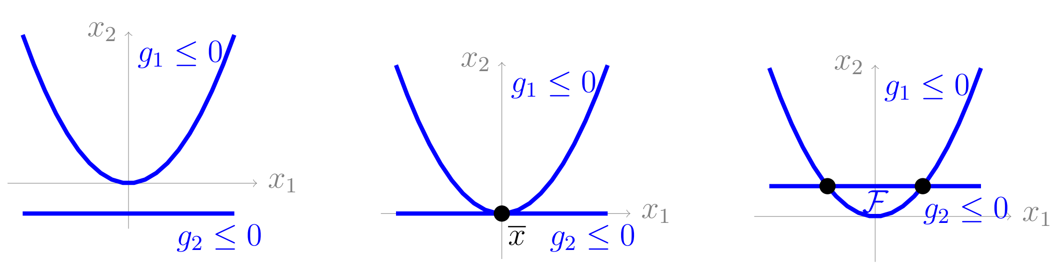

FIGURE 6.45. Picture for: $t < 0$, $\overline{t} = 0$, $0 < t$.

So for $P(t)$, there is a part of the set C_{gc} given locally by $\tilde{C}_{gc} := \{(x, t) \mid x_2 = t,\ x_1^2 = t,\ t \geq 0\}$ (intersection of $g_1 = g_2 = 0$) (see Figure 6.46) with a gc-point of Type 4 at $(\overline{x}, \overline{t}) = (0, 0)$.
Notice that depending on the choice for an objective f, the set C_{gc} could also contain other branches corresponding to critical points where only one (or no) of the constraints is active.

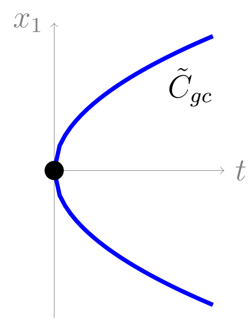

FIGURE 6.46. The set $\tilde{C}_{gc}$ for $P(t)$.

For linear constraints, generically, such Type 4 gc-points are no more possible. To see this consider now a problem $Q(t)$ with the (linear) constraints (see Figure 6.47)

$$g_1(x,t) = x_2 - 1 \le 0 \ , \quad g_2(x,t) = -x_1 t + x_2 \le 0 \ .$$

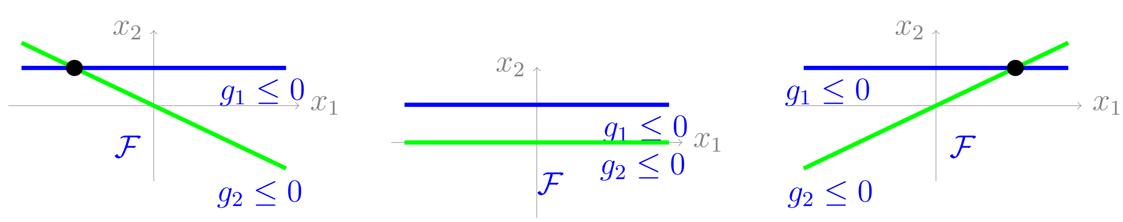

FIGURE 6.47. Picture for: $t < 0$, $\bar{t} = 0$, $0 < t$.

So, independent of f, the set of critical points of $Q(t)$ contains the set

$$\tilde{C}_{gc} = \{(x,t) \mid g_1(x,t) = g_2(x,t) = 0\} = \{(x,t) \mid x = (\frac{1}{t}, 1), \ \ t \neq 0\} \ ,$$

with a two-sided point $\hat{t} = 0$ of Type 4 at infinity (see Figure 6.48).
Note again, that depending on the choice for an objective f the set C_{gc} could also contain other branches corresponding to critical points of $Q(t)$ with only one (or no) active constraint.

For general (nonlinear) programs $P(t)$ a gc-point $(\overline{x}, \overline{t})$ of Type 4 as in Figure 6.45 with two active constraints can be generalized to higher dimensions $n > 2$:

$$g_1(x,t) = \sum_{i=1}^{n-1} x_i^2 \ - x_n \le 0 \ , \quad g_2(x,t) = x_n - t \le 0 \ .$$

Also here at $(\overline{x}, \overline{t}) = (0,0) \in \mathbb{R}^n \times \mathbb{R}$ a generalized critical point of Type 4 occurs.

This is generically excluded for one-parametric programs $Q(t)$, since the constraints are linear. A Type 4 gc-point $(\overline{x}, \overline{t})$ would be a point satisfying with $J_0 := J_{(\overline{x},\overline{t})}$, $|J_{(\overline{x},\overline{t})}| \le n$, the equations

$$B_{J_0}(\overline{t})\overline{x} = b_{J_0}(\overline{t}) \quad \text{and} \quad \text{rank } B_{J_0}(\overline{t}) = k < |J_0| \ . \qquad (6.53)$$

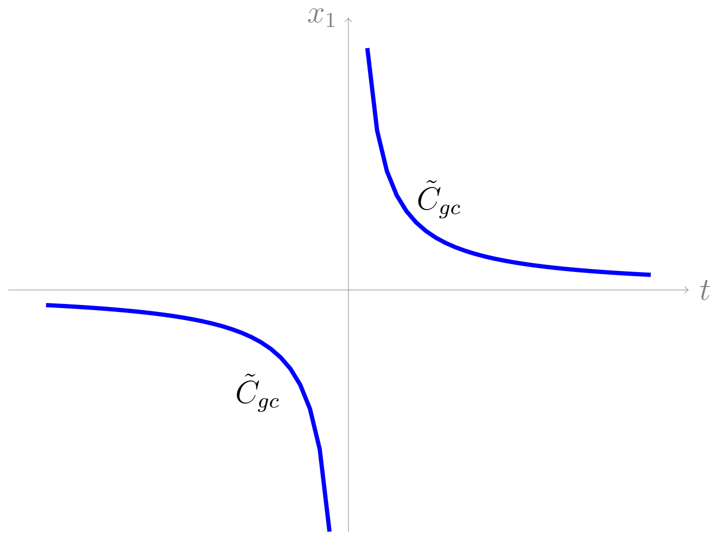

FIGURE 6.48. The set $\tilde{C}_{gc}$ for $Q(t)$.

We show that the existence of such a point is generically excluded. To do so, we consider two possible cases for $(\overline{x}, \overline{t})$:

case $|J_0| < n$: This means that at $\overline{t}$ the $(|J_0| \times n)$-matrix $B_{J_0}(\overline{t})$ meets the set $M^{|J_0|,n}(k)$ of matrices of rank $k < |J_0| < n$. This set is a manifold with codimension (see Theorem 5.11)

$$\text{codim } M^{|J_0|,n}(k) = (n-k)(|J_0| - k) \geq (n-k) \geq 2 \ .$$

However, this situation is generically excluded for a one-parametric matrix function $B_{J_0}(t)$.

case $|J_0| = n$: The second condition in (6.53) means that at $\overline{t}$ the matrix $B_{J_0}(\overline{t})$ meets the set $M^{n,n}(k)$ for $k < n$.
For $k = n-1$ the set $M^{n,n}(n-1)$ has codimension $(n-(n-1))^2 = 1$. So, this situation may happen in the generic situation (on a discrete set of points $\overline{t} \in \mathbb{R}$). But the first condition in (6.53) implies ($|J_0| = n$)

$$\text{rank } (B_{J_0}(\overline{t}), b_{J_0}(\overline{t})) = \text{rank } B_{J_0}(\overline{t}) < n \ .$$

So, the matrix $(B_{J_0}(\overline{t}), b_{J_0}(\overline{t})) \in M^{n,n+1}$ has rank $< n$, i.e., this matrix meets the set $M^{n,n+1}(k)$ with $k < n$. But $M^{n,n+1}(k)$ then satisfies

$$\text{codim } M^{n,n+1}(k) = (n+1-k)(n-k) \geq 2 \ ,$$

and consequently also this situation is generically excluded. In other words, even if at $\overline{t}$ the n gradients $\nabla_x g_j(\overline{x}, \overline{t}) = B_{j\cdot}(\overline{t})$, $j \in J_0$, $|J_0| = n$, are linearly dependent, an intersection point $\overline{x}$ of the n constraints

$$B_{j\cdot}(\overline{t})x = b_j(\overline{t}) \ , \quad j \in J_0 \ ,$$

is generically excluded.

6.6.2. Genericity results for one-parametric linear programs.
We finally examine the properties of one-parametric linear programs in primal form (cf., Section 3.2),

$$L(t): \quad \min\ c(t)^T x \quad \text{s.t.} \quad x \in \mathcal{F}(t) = \{x \in \mathbb{R}^n \mid B(t)x \le b(t)\}\ , \tag{6.54}$$

depending on the parameter $t \in \mathbb{R}$, where $c \in C^\infty(\mathbb{R}, \mathbb{R}^n)$, $B \in C^\infty(\mathbb{R}, M^{m,n})$ $b \in C^\infty(\mathbb{R}, \mathbb{R}^m)$. We refer the reader also to [11].

In order to facilitate a direct comparison with the results for quadratic programs $Q(t)$ in Subsection 6.6.1 and for general nonlinear programs $P(t)$ in Subsection 6.5.2, here, in contrast to Section 3.2, the primal program $L(t)$ is formulated as a minimum problem.

For the one-parametric problem $L(t)$ above (n, m fixed), the problem set is given by

$$\mathcal{P} = \{(c, B, b) \in C^\infty(\mathbb{R}, \mathbb{R}^n \times M^{m,n} \times \mathbb{R}^m)\} \equiv C^\infty(\mathbb{R}, \mathbb{R}^N\}\ ,\ N = n + n \cdot m + m\ , \tag{6.55}$$

and we often write $L(c, B, b)$ instead of $L(t)$.

For $L(t)$ a feasible point $(\overline{x}, \overline{t})$, $\overline{x} \in \mathcal{F}(\overline{t})$, is a gc-point if with $f(x,t) := c(t)^T x$, $g_j(x,t) := B_{j\cdot}(t)x - b_j(t)$, the gradients

$$\nabla_x^T f(\overline{x}, \overline{t}) = c(\overline{t})\ ,\quad \nabla_x^T g_j(\overline{x}, \overline{t}) = B_{j\cdot}(\overline{t})^T,\ j \in J_{(\overline{x},\overline{t})}\ ,\ \text{ are linearly dependent.} \tag{6.56}$$

LICQ holds at $(\overline{x}, \overline{t})$ if

$$b_{j\cdot}(\overline{t})^T\ ,\quad j \in J_{(\overline{x},\overline{t})}\ ,\ \text{ are linearly independent}\ , \tag{6.57}$$

and such a point $(\overline{x}, \overline{t})$ is a critical point for $L(t)$ if LICQ holds as well as the critical point relation

$$c(\overline{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \overline{\mu}_j B_{j\cdot}(\overline{t})^T = 0 \quad \text{with (unique) multipliers } \overline{\mu}_j \in \mathbb{R}\ . \tag{6.58}$$

Again we say that

$$(SC) \quad \text{holds, if in (6.58) we have:} \quad \overline{\mu}_j \ne 0\ \forall j \in J_{(\overline{x},\overline{t})}\ . \tag{6.59}$$

As before we are interested in the generic structure of the set C_{gc} of gc-points of $L(t)$ and the set C_m corresponding to (global) minimizers. More precisely, we wish to know which types of gc-points can generically occur for programs $L(t)$.
As we shall see there are differences between the generic behaviour of these one-parametric linear problems $L(t)$ and the properties of the quadratic programs $Q(t)$ (in Subsection 6.6.1) and the general nonlinear programs $P(t)$ (in Subsection 6.5.2). We start with three differences concerning gc-points of Tye 1,2 and 3.

Firstly consider a gc-point $(\overline{x}, \overline{t})$ of Type 1 for $L(t)$. The critical point equations and the feasibility conditions for (x, t) read:

$$\begin{array}{rcl} \sum_{j \in J_{(\overline{x},\overline{t})}} \mu_j B_{j\cdot}(t)^T &=& -c(t) \\ B_{j\cdot}(t)x &=& b_j(t)\ ,\quad j \in J_{(\overline{x},\overline{t})} \end{array}\ . \tag{6.60}$$

In contrast to the corresponding equations for $Q(t)$ or $P(t)$, these relations (6.60) in the unknowns (x, μ) are completely decoupled. For a Type 1 critical point $(\overline{x}, \overline{t})$ for $L(t)$

we in particular require that at $\bar{t}$ the system (6.60) has a unique solution $(\overline{x}, \overline{\mu})$, i.e., the $(n + |J_{(\overline{x},\bar{t})}|) \times (n + |J_{(\overline{x},\bar{t})}|)$-matrix

(6.61) $$\begin{pmatrix} 0 & B^T_{J_{(\overline{x},\bar{t})}}(\bar{t}) \\ B_{J_{(\overline{x},\bar{t})}}(\bar{t}) & 0 \end{pmatrix} \quad \text{is nonsingular .}$$

This obviously implies rank $B_{J_{(\overline{x},\bar{t})}}(\bar{t}) = |J_{(\overline{x},\bar{t})}|$ and rank $B_{J_{(\overline{x},\bar{t})}}(\bar{t}) = n$, which is equivalent to

(6.62) $$|J_{(\overline{x},\bar{t})}| = n \quad \text{and} \quad \text{rank } B_{J_{(\overline{x},\bar{t})}}(\bar{t}) = n \; .$$

Consequently, for a Type 1 point $(\overline{x}, \bar{t})$ of $L(t)$ the point $\overline{x}$ is always a non-degenerate vertex of $\mathcal{F}(\bar{t})$.

Secondly, it is evident, that since the Hessians of all problem functions in a linear program is identical zero, for $L(t)$, no gc-points $(\overline{x}, \bar{t})$ of Type 3 as for $P(t)$, or points $\hat{t}$ of Type 3 at infinity as for $Q(t)$ can occur.

Thirdly, generically the gc-points $(\overline{x}, \bar{t})$ of $P(\bar{t})$, $Q(\bar{t})$ are always locally unique (isolated). Consider, e.g., the sketches of the feasible sets (in blue) and level sets (in red) of a quadratic program $Q(t)$ for $x \in \mathbb{R}^2,\ t \in \mathbb{R}$ (see Figure 6.49, picture for $t_1 < \bar{t} < t_2$). We have given 3 linear constraints $g_j(x,t) := B_{j\cdot}(t)x - b_j(t) \le 0,\ j = 1, 2, 3$.

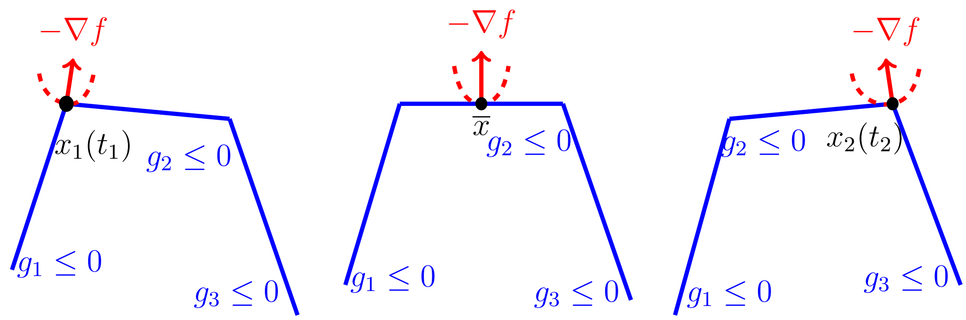

FIGURE 6.49. Picture for $t = t_1 < \bar{t}$, $t = \bar{t}$, $t = t_2 > \bar{t}$.

The intersection points $x_1(t)$ of $g_1 = g_2 = 0$, and $x_2(t)$ of $g_2 = g_3 = 0$, are always critical points, and obviously, at $(x_1(t_1), t_1)$ and $(x_2(t_2), t_2)$ we have given a critical point of Type 2, and at $(\overline{x}, \bar{t})$ a point of Type 1 for $Q(t)$. The corresponding set C_{gc} is sketched in Figure 6.50.

For the linear program $L(t)$ with the same linear constraints sketched in Figure 6.51 the situation is essentially different.

Here, for $t = t_1$ ($t = t_2$) the point $x_1(t_1)$ ($x_2(t_2)$) is a unique (global) vertex minimizer of Type 1 for $L(t_1)$ ($L(t_2)$). For any $t \ne \bar{t}$ near $\bar{t}$ the intersection points $x_1(t)$ of $g_1 = g_2 = 0$ and $x_2(t)$ of $g_2 = g_3 = 0$ are locally unique critical points (primal vertices) of $L(t)$. But at $t = \bar{t}$ the whole line segment

$$S(\bar{t}) = \{(1-\tau)x_1(\bar{t}) + \tau x_2(\bar{t}) \mid \tau \in [0,1]\}$$

consist of critical points. So, for $t \approx \bar{t}$ the set C_{gc} looks as given in Figure 6.52.

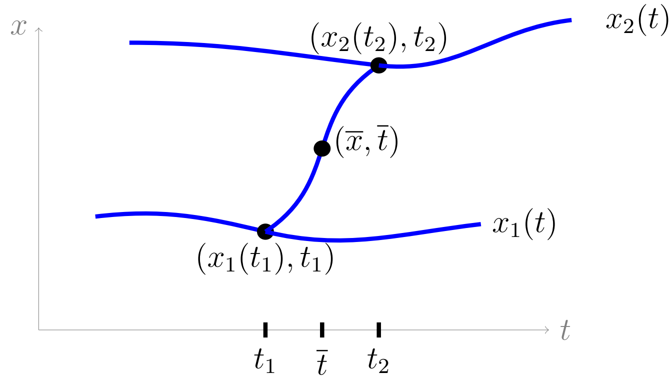

FIGURE 6.50. Set C_{gc} of $Q(t)$ with two points $(x_1(t_1), t_1)$, $(x_2(t_2), t_2)$ of Type 2.

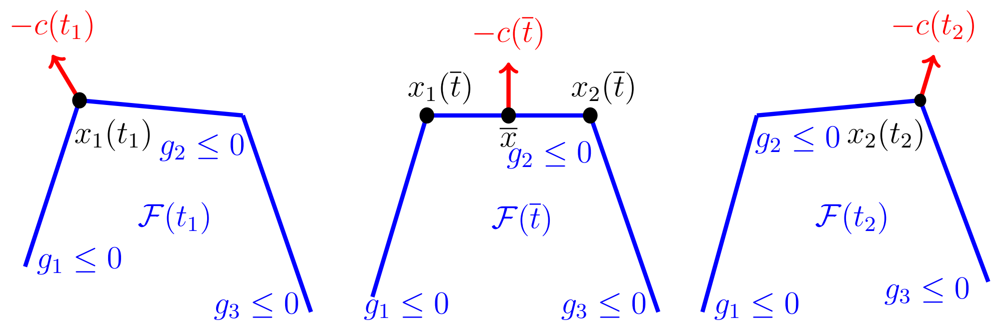

FIGURE 6.51. Picture for $t = t_1 < \overline{t}$, $t = \overline{t}$, $t = t_2 > \overline{t}$.

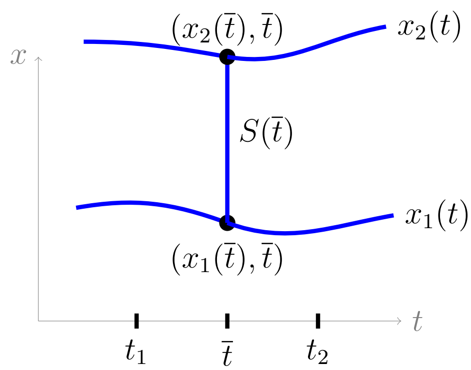

FIGURE 6.52. Set C_{gc} of $L(t)$ containing a whole line segment $S(\overline{t})$ of gc-points of $L(\overline{t})$.

The end points $(x_1(\overline{t}), \overline{t})$ and $(x_2(\overline{t}), \overline{t})$ of $S(\overline{t})$ will be called gc-points $(\overline{x}, \overline{t})$ of Type 2^* of $L(t)$.
In the sequel we will show that generically for one-parametric programs $L(t)$ only gc-points of Type 1, 2^* , 5 and points $\hat{t}$ of Type 4 at infinity can occur.

Recall that a problem $L(t) = L(c, B, b)$ is given by an element $(c, B, b) \in \mathcal{P} = C^\infty(\mathbb{R}, \mathbb{R}^n \times M^{m,n} \times \mathbb{R}^m)$ (cf., (6.55)). For an index set $J_0 \subset J$ we again define

$$B_{J_0}(t) = \begin{pmatrix} B_{j\cdot}(t) \\ \vdots \end{pmatrix}_{j \in J_0} \in M^{|J_0|,n} , \quad b_{J_0}(t) = \begin{pmatrix} b_j(t) \\ \vdots \end{pmatrix}_{j \in J_0} \in \mathbb{R}^{|J_0|} .$$

Then a gc-point $(\overline{x}, \overline{t})$ of Type 1, 2^* or 5 and a point $\overline{t}$ of Type 4 at infinity will be a point where the vector

$$\big(c(\overline{t}), B_{J_0}(\overline{t}), b_{J_0}(\overline{t})\big) \quad \text{with some } J_0 \subset J$$

meets a certain manifold in $(\mathbb{R}^n \times M^{|J_0|,n} \times \mathbb{R}^m)$. We now describe the different types of gc-points which can appear for $L(t)$ in the generic case.

Type 1: A point $(\overline{x}, \overline{t})$, $\overline{x} \in \mathcal{F}(\overline{t})$, where LICQ holds with $|J_{(\overline{x},\overline{t})}| = n$ and the critical point equation

$$c(\overline{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \overline{\mu}_j B_{j\cdot}(\overline{t})^T = 0$$

is valid with SC: $\overline{\mu}_j \neq 0, \ j \in J_{(\overline{x},\overline{t})}$.
Under this condition, the point $\overline{x}$ is a nondegenerate primal vertex of $\mathcal{F}(\overline{t})$ (cf., Definition 3.1) and $\overline{\mu}$ is a non-degenerate dual vertex.

Type 5: A point $(\overline{x}, \overline{t})$, $\overline{x} \in \mathcal{F}(\overline{t})$, satisfying

(5a) $|J_{(\overline{x},\overline{t})}| = n + 1$ (so LICQ fails)

(5b)

$$\nabla^T_{(x,t)} g_j(\overline{x}, \overline{t}) = \begin{pmatrix} B_{j\cdot}(\overline{t})^T \\ B'_{j\cdot}(\overline{t})\overline{x} - b'_j(\overline{t}) \end{pmatrix} , \quad j \in J_{(\overline{x},\overline{t})} , \ \text{are linearly independent} ,$$

where $B'_{j\cdot}(\overline{t}) = \frac{d}{dt} B_{j\cdot}(\overline{t}), \ b'_j(\overline{t}) = \frac{d}{dt} b_j(\overline{t})$.

(5c) In view of (5a), (5b) there is (up to a common factor) a unique nonzero solution $\mu \in \mathbb{R}^{|J_{(\overline{x},\overline{t})}|}$ of

$$\sum_{j \in J_{(\overline{x},\overline{t})}} \mu_j B_{j\cdot}(\overline{t})^T = 0 . \tag{6.63}$$

We have: $\mu_j \neq 0, \ \forall j \in J_{(\overline{x},\overline{t})}$.

(5d) By (5a), (5b) there exist unique $y_j, \ j \in J_{(\overline{x},\overline{t})}$, solving

$$\nabla^T_{(x,t)} f(\overline{x}, \overline{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} y_j \nabla^T_{(x,t)} g_j(\overline{x}, \overline{t}) = 0 ,$$

or,

$$\begin{pmatrix} c(\overline{t}) \\ c'(\overline{t})\overline{x} \end{pmatrix} + \sum_{j \in J_{(\overline{x},\overline{t})}} y_j \begin{pmatrix} B_{j\cdot}(\overline{t})^T \\ B'_{j\cdot}(\overline{t})\overline{x} - b'_j(\overline{t}) \end{pmatrix} = 0 . \tag{6.64}$$

Defining $\gamma_{qj} := y_j - y_q \frac{\mu_j}{\mu_q}, \ j, q \in J_{(\overline{x},\overline{t})}$, we have:

$$\gamma_{q,j} \neq 0 \qquad \forall q \neq j, \ q, j \in J_{(\overline{x},\overline{t})} .$$

Type 2*: A gc-point $(\hat{x}, \hat{t})$ such that we have:

(2a) $|J_{(\hat{x},\hat{t})}| = n$ and rank $B_{J_{(\hat{x},\hat{t})}}(\hat{t}) = n$,
i.e., $\hat{x}$ is a non-degenerate vertex of $\mathcal{F}(\hat{t})$.

(2b) By (2a) there are unique multipliers $\mu_j \in \mathbb{R},\ j \in J_{(\hat{x},\hat{t})}$, such that

$$c(\hat{t}) + \sum_{j \in J_{(\hat{x},\hat{t})}} \mu_j B_{j\cdot}(\hat{t})^T = 0 \ .$$

For precisely one $j_0 \in J_{(\hat{x},\hat{t})}$ we have: $\mu_{j_0} = 0$.

(2c) By (2a) there is (up to a positive factor) a unique solution $d \neq 0$ of

$$\begin{array}{ll} B_{J_{(\hat{x},\hat{t})}\setminus\{j_0\}}(\hat{t})d & = 0 \\ B_{j_0\cdot}(\hat{t})d & < 0 \end{array} \ . \tag{6.65}$$

If we define

$$\tau_0 := \min_{\substack{j \in J \setminus J_{(\hat{x},\hat{t})} \\ B_{j\cdot}(\hat{t})d > 0}} \frac{b_j(\hat{t}) - B_{j\cdot}(\hat{t})\hat{x}}{B_{j\cdot}(\hat{t})d} \ > \ 0 \ ,$$

then in case $\tau_0 = \infty$ the set

$$S(\hat{t}) = \{\hat{x} + \tau d \mid 0 \leq \tau \leq \infty\} \ , \qquad \text{half-line,}$$

is a set of critical points of $L(\hat{t})$. In case $\tau_0 < \infty$ the minimum value τ_0 is attained at precisely one index $j_1 \in J \setminus J_{(\hat{x},\hat{t})}$ and the set

$$S(\hat{t}) = \{\hat{x} + \tau d \mid 0 \leq \tau \leq \tau_0\} \ , \quad \text{line-segment,}$$

is a set of critical points of $L(\hat{t})$. Moreover, $B_{J_{(\hat{x},\hat{t})}\setminus\{j_0\}\cup\{j_1\}}(\hat{t})$ has rank n and the point $\tilde{x} := \hat{x} + \tau_0 d$ is a nondegenerate vertex of $\mathcal{F}(\hat{t})$ with $J_{(\tilde{x},\hat{t})} = J_{(\hat{x},\hat{t})} \setminus \{j_0\} \cup \{j_1\}$.

To characterize a point of Type 4 at infinity, as for the program $Q(t)$ in Subsection 6.6.1, we consider a critical point $(\overline{x}, \overline{t})$ of Type 1 on a maximal component of the one-dimensional manifold of Type 1 points, parameterised by the curve $(x(t), t),\ t \in I_{(\overline{x},\overline{t})}$, with a maximal open interval $I_{(\overline{x},\overline{t})} \subset \mathbb{R}$ containing $\overline{t}$. We define (see (6.52))

$$\hat{t} := \sup \{t \in I_{(\overline{x},\overline{t})}\} \qquad \big(\text{or} \quad \hat{t} := \inf \{t \in I_{(\overline{x},\overline{t})}\}\big) \ . \tag{6.66}$$

For $\hat{t} = \infty$ (or $\hat{t} = -\infty$) the picture is clear.

Type 4 at infinity: A point $\hat{t} \in \mathbb{R}$ defined by (6.66) (corresponding to the Type 1 point $(\overline{x}, \overline{t})$ and the curve $(x(t), t)$ above) which satisfies the following:

(4a) $|J_0| = n$ where $J_0 := J_{(\overline{x},\overline{t})}$

(4b) rank $B_{J_0}(\hat{t}) = n - 1$ and rank $\big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) = n$.

(4c)

$$\lim_{t \uparrow \hat{t}} \frac{x(t)}{\|x(t)\|} = \tilde{x} \ ,$$

where $\tilde{x}$ is (up to a sign) the unique vector satisfying

$$B_{J_0}(\hat{t})\tilde{x} = 0 \;, \qquad \|\tilde{x}\| = 1 \;.$$

We now formulate the genericity result for one-parametric linear programs of the type $L(t) = L(c, B, b)$ similar to Theorem 6.12. This result will be based on the Jet Transversality Theorem 6.1. More precisely, we show that a generic subset of problems $(c, B, b) \in \mathcal{P} \equiv C^\infty(\mathbb{R}, \mathbb{R}^n \times M^{m,n} \times \mathbb{R}^m)$ avoid certain manifolds (of codimension 2) of $\mathbb{R}^n \times M^{m,n} \times R^m$ and only meet other manifolds (of codimension 1) at discrete point sets. Since these manifolds are not closed we will not obtain an "open and dense" result but only a "genericity" result, see Theorem 6.1.

THEOREM 6.14. [Genericity result for one-parametric linear programs]
There is a generic subset $\mathcal{P}_r$ of $\mathcal{P}$ such that for each $(c, B, b) \in \mathcal{P}_r$ the following holds: For any point $\hat{t}$ defined in (6.66) *precisely one of the following alternatives (a) or (b) holds for $L(c, B, b)$.*

(a) *We have*

$$\lim_{t \uparrow \hat{t}} x(t) = \hat{x} \qquad \textit{for some } \hat{x} \in \mathbb{R} \;, \tag{6.67}$$

and then either

(i) *$(\hat{x}, \hat{t})$ is a gc-point of Type 5 and $|J_{(\hat{x},\hat{t})}| = n + 1$, or*

(ii) *$(\hat{x}, \hat{t})$ is a point of Type 2* and $|J_{(\hat{x},\hat{t})}| = n$.*

(b) *We have*

$$\lim_{t \uparrow \hat{t}} \|x(t)\| = \infty \;, \tag{6.68}$$

and then $\hat{t}$ is a point of Type 4 at infinity.

The set of points $\hat{t} \in \mathbb{R}$ in (a) and (b) is a discrete set.

Proof. (a) Assuming (6.67), by continuity, the point $(\hat{x}, \hat{t})$ must be a gc-point. More precisely, with some $\mu_0 \in \mathbb{R}$, $\mu_j \in \mathbb{R},\ j \in J_{(\overline{x},\overline{t})}$, $|J_{(\overline{x},\overline{t})}| = n$, it holds

$$\mu_0 c(\hat{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \mu_j B_{j\cdot}(\hat{t})^T \;=\; 0 \;, \qquad |\mu_0| + \sum_{j \in J_{(\overline{x},\overline{t})}} |\mu_j| = 1 \tag{6.69}$$

$$B_{J_{(\overline{x},\overline{t})}}(\hat{t})\hat{x} \;=\; b_{J_{(\overline{x},\overline{t})}}(\hat{t}) \;. \tag{6.70}$$

The following cases may occur. Note that by assumption, $(\hat{x}, \hat{t})$ is not a point of Type 1.

(a1) rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n$, and $J_{(\hat{x},\hat{t})} = J_{(\overline{x},\overline{t})} \cup \{j_0\}$. Then as we shall see, generically, $(\hat{x}, \hat{t})$ is a point of Type 5.

(a2) rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n$, and $J_{(\hat{x},\hat{t})} = J_{(\overline{x},\overline{t})} \cup J_1$, $|J_1| \geq 2$. So putting $J_0 := J_{(\hat{x},\hat{t})}$ the matrix

$$\big(B_{J_0}(\hat{t})\; b_{J_0}(\hat{t})\big) \in M^{n+q,n+1}, q \geq 2 \;, \quad \text{has rank } \leq n \;,$$

i.e., $\big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big)$ meets a manifold of codimension $\geq (n+2-n)(n+1-n) = 2$ (see Theorem 5.11). This can be excluded for a generic subset $\mathcal{P}_{a2}$ of $\mathcal{P}$.

(a3) rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = k < n$, and $J_{(\hat{x},\hat{t})} = J_{(\overline{x},\overline{t})} \cup J_1$, $|J_1| \geq 0$. Together with (6.70) this condition give just the relations (6.53) which have been shown to be excluded for a generic subset $\mathcal{P}_{a3}$ of $\mathcal{P}$ (with arguments after (6.53) in Subsection 6.6.1).
(a4) rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n$, $J_{(\hat{x},\hat{t})} = J_{(\overline{x},\overline{t})}$, so that

$$c(\hat{t}) + \sum_{j \in J_{(\overline{x},\overline{t})}} \mu_j B_{j\cdot}(\hat{t})^T = 0 \quad \text{with unique } \mu_j \in \mathbb{R}\ . \tag{6.71}$$

In this case (since $(\hat{x},\hat{t})$ is not of Type 1) we can assume that $q \geq 1$ of these μ_js are equal to zero.
For $q = 1$, as we shall show below, $(\hat{x},\hat{t})$ is a point of Type 2*. But the case $q \geq 2$ is generically excluded. Indeed for $q \geq 2$ the condition (6.71) means that for a subset $J_0 \subset J_{(\overline{x},\overline{t})}$, $|J_0| = |J_{(\overline{x},\overline{t})}| - q \leq n-2$ ($\mu_j = 0$ for $j \in J_{(\overline{x},\overline{t})} \setminus J_0$) the $n \times (|J_0|+1)$-matrix

$$\big(B_{J_0}(\hat{t})^T\ c(\hat{t})\big) \quad \text{has rank } |J_0|\ ,$$

i.e., this matrix meets a manifold of codimension $(n-|J_0|)(|J_0|+1-|J_0|) = (n-|J_0|) \geq 2$ (cf., Theorem 5.11). Again this is excluded for a generic subset $\mathcal{P}_{a4}$ of $\mathcal{P}$.

Proof of (b): Let (6.68) hold. By assumption, the Type 1 points $(x(t),t)$ satisfy with $|J_{(\overline{x},\overline{t})}| = n$:

$$B_{J_{(\overline{x},\overline{t})}}(t)x(t) = b_{J_{(\overline{x},\overline{t})}}(t) \quad \forall t < \hat{t},\ t \in I(\overline{x},\overline{t})\ .$$

By dividing this relation by $\|x(t)\|$ using (6.68) we find that $\frac{x(t)}{\|x(t)\|}$, $t \uparrow \hat{t}$, has some limit point $\tilde{x}$, $\|\tilde{x}\| = 1$ satisfying

$$B_{J_{(\overline{x},\overline{t})}}(\hat{t})\tilde{x} = 0\ . \tag{6.72}$$

So rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = q \leq n-1$. Again for a generic subset $\mathcal{P}_{b1}$ of $\mathcal{P}$ the condition $q \leq n-2$ is excluded and the case $q = n-1$ is possible for a discrete set of points $\hat{t} \in \mathbb{R}$.
So let $q = n-1$. Since rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n-1$, the relation $B_{J_{(\overline{x},\overline{t})}}(\hat{t})\tilde{x} = 0$ can only have (up to a sign) one solution $\tilde{x}$, $\|\tilde{x}\| = 1$. Consequently $\frac{x(t)}{\|x(t)\|}$, $t \uparrow \hat{t}$, can only have $\tilde{x}$ as a limit point (up to a sign), i.e.,

$$\lim_{t \uparrow \hat{t}} \frac{x(t)}{\|x(t)\|} = \tilde{x}\ . \tag{6.73}$$

Moreover, for a generic subset $\mathcal{P}_{b2}$ of $\mathcal{P}$ we have rank $\big(B_{J_{(\overline{x},\overline{t})}}(\hat{t})\ b_{J_{(\overline{x},\overline{t})}}(\hat{t})\big) = n$. Indeed the case that the $n \times (n+1)$-matrix $\big(B_{J_{(\overline{x},\overline{t})}}(\hat{t})\ b_{J_{(\overline{x},\overline{t})}}(\hat{t})\big)$ has rank $\leq n-1$ (meets a manifold of codimension ≥ 2) can be generically excluded.

Proof of (a) continued: We now analyse further the remaining cases (a1) and (a4) in (a).

case (a1), a point $(\hat{x},\hat{t})$ of Type 5:

By assumption, the point $(\hat{x}, \hat{t})$ satisfies with $J_{(\overline{x},\overline{t})}$, $|J_{(\overline{x},\overline{t})}| = n$,

$$\text{rank } B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n \quad \text{and} \quad J_{(\hat{x},\hat{t})} = J_{(\overline{x},\overline{t})} \cup \{j_0\} \ .$$

So $|J_{(\hat{x},\hat{t})}| = n+1$ and condition (5a) of Type 5 is valid. We now prove the other conditions (5b)-(5d).

(5b): Defining $J_0 := J_{(\hat{x},\hat{t})}$ then $|J_0| = n+1$. Note that by the assumption rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n$, the matrix

$$\big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) \in M^{n+1} \quad \text{has rank } n \ .$$

So, in view of Ex. 5.5 this condition is equivalent with

$$\det\big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) = 0 \ ,$$

or using the feasibility condition $g_{J_0}(\hat{x}, \hat{t}) = B_{J_0}(\hat{t})\hat{x} - b_{J_0}(\hat{t}) = 0$, it is also equivalent with

$$\det\big(B_{J_0}(\hat{t})\ g_{J_0}(\hat{x}, \hat{t})\big) = 0 \ .$$

We now apply Ex. 5.9 with $S = M^{n+1}(n)$, $f(t) := \big(B_{J_0}(t)\ \ g_{J_0}(\hat{x}, t)\big)$ and $h(M) = \det(M)$, $M \in M^{n+1}$. By Theorem 6.1 there is a generic subset $\mathcal{P}_{51}$ of $\mathcal{P}$ such that for $(c, B, b) \in \mathcal{P}_{51}$ we have $f \,\overline{\pitchfork}\, M^{n+1}(n)$, or, equivalently (cf., Ex. 5.9)

$$0 \neq \frac{d}{dt} h(f(\hat{t})) = \frac{d}{dt} \det\big(B_{J_0}(\hat{t})\ g_{J_0}(\hat{x}, \hat{t})\big) = \det\Big(B_{J_0}(\hat{t})\ \frac{d}{dt} g_{J_0}(\hat{x}, \hat{t})\Big) \ ,$$

where the last equality follows by using the multi-linearity of the determinant and $g_{J_0}(\hat{x}, \hat{t}) = 0$. This relation implies that the $n + 1$ rows of $\big(B_{J_0}(\hat{t})\ \frac{d}{dt} g_{J_0}(\hat{x}, \hat{t})\big)$, given by

$$\nabla_{(x,t)} g_j(\hat{x}, \hat{t}) = \big(B_{j\cdot}(\hat{t})\ ,\ B'_{j\cdot}(\hat{t})\hat{x} - b'_{j\cdot}(\hat{t})\big)\ ,\quad j \in J_0 = J_{(\hat{x},\hat{t})}, \quad \text{are linearly independent} \ .$$

(5c): Recall that by the assumption rank $B_{J_{(\overline{x},\overline{t})}}(\hat{t}) = n$, $J_0 = J_{(\hat{x},\hat{t})}$, $|J_0| = n + 1$, there exist (up to a factor) a unique nonzero multiplier μ_{J_0}, such that

$$\sum_{j \in J_0} \mu_j B_{j\cdot}(\hat{t})^T = 0 \ . \tag{6.74}$$

We now show that the condition $\mu_{j_1} = 0$ for some $j_1 \in J_0$ can be generically excluded. To do so, note that by assumption there exist unique multipliers $\gamma_j,\ j \in J_{(\overline{x},\overline{t})}$, such that

$$\sum_{j \in J_{(\overline{x},\overline{t})}} \gamma_j B_{j\cdot}(\hat{t})^T = -B_{j_0\cdot}(\hat{t})^T \ .$$

If in (6.74) we would have $\mu_{j_1} = 0$, then $j_1 \in J_{(\overline{x},\overline{t})}$ and there would exist a solution $\gamma \in \mathbb{R}^{n-1}$ of

$$B_{J_{(\overline{x},\overline{t})} \setminus \{j_1\}}(\hat{t})^T \gamma = -B_{j_0\cdot}(\hat{t})^T \ .$$

Thus, together with the feasibility condition $B_{J_0}(\hat{t})\hat{x} = b_{J_0}(\hat{t})$ (recall $J_0 = J_{(\overline{x},\overline{t})} \cup \{j_0\}$) there must be a solution $(\gamma, x) \in \mathbb{R}^{n-1} \times \mathbb{R}^n$ of

$$\begin{array}{lll} B_{J_{(\overline{x},\overline{t})} \setminus \{j_1\}}(\hat{t})^T \gamma & & = -B_{j_0\cdot}(\hat{t})^T \\ & B_{J_{(\overline{x},\overline{t})}}(\hat{t}) x & = b_{J_{(\overline{x},\overline{t})}}(\hat{t}) \\ & B_{j_0\cdot}(\hat{t}) x & = b_{j_0}(\hat{t}) \end{array} \ . \tag{6.75}$$

Since rank $B_{J_{(\overline{x},\overline{t})}\setminus\{j_1\}}(\hat{t}) = n-1$ (after appropriate operations of the first n rows) we can assume that this system has the form:

$$\begin{array}{llll} A_{11}\gamma & & = -c_1 & A_{11} \in M^{n-1} \text{ is nonsingular} \\ a_0^T\gamma & & = -c_0 & a_0, c_1 \in \mathbb{R}^{n-1}, c_0 \in \mathbb{R} \\ & B_{11}x & = b_1 & B_{11} \in M^n \text{ is nonsingular} \\ & b_0^T x & = \gamma_0 & b_0, b_1 \in \mathbb{R}^n, \gamma_0 \in \mathbb{R} \end{array} .$$

Solving this system using $\gamma = -A_{11}^{-1}c_1$, $x = B_{11}^{-1}b_1$ leads to the relations

$$c_0 - a_0^T A_{11}^{-1} c_1 = 0 \ , \quad \gamma_0 - b_0^T B_{11}^{-1} b_1 = 0 \ ,$$

which obviously define two linearly independent equations in the unknowns $(A_{11}, a_0, c_1, c_0, B_{11}, b_0, b_1, \gamma_0)$. Consequently, the condition $\mu_{j_1} = 1$ means that the problem (c, B, b) at $\hat{t}$ meets a manifold of codimension 2, which can generically be avoided.

<u>(5d):</u> By (5b) there exist unique solutions $y_j \in \mathbb{R}, \ j \in J_0$, of

$$\begin{pmatrix} c(\hat{t}) \\ c'(\hat{t})\hat{x} \end{pmatrix} + \sum_{j \in J_0} y_j \begin{pmatrix} B_{j\cdot}(\hat{t})^T \\ B'_{j\cdot}(\hat{t})\hat{x} - b'_j(\hat{t}) \end{pmatrix} = 0 \ . \tag{6.76}$$

For each $q \in J_0$ we subtract $y_q \frac{1}{\mu_q}$ times (6.74) from the first n rows of (6.76) to obtain

$$c(\hat{t}) + \sum_{j \in J_0\setminus\{q\}} \big(y_j - y_q \frac{\mu_j}{\mu_q}\big) B_{j\cdot}(\hat{t})^T = 0 \ , \quad q \in J_0 \ .$$

Assume now $\gamma_{q_1,j_1} = 0$ holds for some pair $(q_1, j_1), \ q_1 \neq j_1$ (recall $\gamma_{q,j} = y_j - y_q \frac{\mu_j}{\mu_q}$), then

$$c(\hat{t}) + \sum_{j \in J_{j \in J_0}\setminus\{q_1,j_1\}} \big(y_j - y_{q_1} \frac{\mu_j}{\mu_{q_1}}\big) B_{j\cdot}(\hat{t})^T = 0 \ .$$

Together with the feasibility conditions (recall $J_0 = J_{(\overline{x},\overline{t})} \cup \{j_0\}$),

$$\begin{array}{ll} B_{J_{(\overline{x},\overline{t})}}(\hat{t})\hat{x} & = b_{J_{(\overline{x},\overline{t})}}(\hat{t}) \\ B_{j_0\cdot}(\hat{t})\hat{x} & \ \ b_{j_0}(\hat{t}) \end{array} ,$$

this means that with $J_1 := J_0 \setminus \{q_1, j_1\}$, $|J_1| = n-1$, there is a solution $(\gamma, x) \in R^{n-1} \times \mathbb{R}^n$ of

$$\begin{array}{lll} B_{J_1}(\hat{t})^T\gamma & & = -c(\hat{t}) \\ & B_{J_{(\overline{x},\overline{t})}}(\hat{t})x & = b_{J_{(\overline{x},\overline{t})}}(\hat{t}) \\ & B_{j_0\cdot}(\hat{t})x & = b_{j_0}(\hat{t}) \end{array} . \tag{6.77}$$

Generically we can assume that the $n \times (n-1)$-matrix $B_{J_1}(\hat{t})^T$ has rank $n-1$ (use again Theorem 5.11). So, (6.77) has the same structure as the system (6.75), and with the same arguments as above in the proof of part (5c), we conclude that the condition $\gamma_{q_1,j_1} = 0$ leads to the condition that the problem (c, B, b) meets at $\hat{t}$ a manifold of codimension 2, which again can generically be excluded.

<u>case (a4), gc-point $(\hat{x}, \hat{t})$ of Type 2^*.</u>

Here, with $J_0 := J_{(\overline{x},\overline{t})} = J_{(\hat{x},\hat{t})}$, $|J_0| = n$, we have rank $B_{J_0}(\hat{t}) = n$, and with unique $\mu_j,\ j \in J_0$,

$$c(\hat{t}) + \sum_{j\in J_0} \mu_j B_{j\cdot}(\hat{t})^T = 0 \ , \ \text{ and } \mu_{j_0} = 0 \text{ for precisely one } j_0 \in J_0 \ . \tag{6.78}$$

Moreover, $\hat{x}$ is a feasible nondegenerate vertex of $\mathcal{F}(\hat{t})$,

$$B_{J_0}(\hat{t})\hat{x} = b_{J_0}(\hat{t}) \ .$$

Now consider the unique solution $d \neq 0$ (unique up to a factor) of

$$B_{J_0\setminus\{j_0\}}(\hat{t})d = 0 \ .$$

We can assume $B_{j_0\cdot}(\hat{t})d < 0$ (choosing the appropriate sign of d). Then obviously, using $J_{(\hat{x},\hat{t})} = J_0$, for the points $x_\tau; = \hat{x} = \tau d$ we find,

$$\begin{array}{ll}
B_{J_0\setminus\{j_0\}}(\hat{t})x_\tau - b_{J_0\setminus\{j_0\}}(\hat{t}) = 0 & \forall \tau > 0 \\
B_{j_0\cdot}(\hat{t})x_\tau - b_{j_0}(\hat{t}) = \tau B_{j_0\cdot}(\hat{t})d < 0 & \forall \tau > 0 \\
B_{j\cdot}(\hat{t})x_\tau - b_j(\hat{t}) = B_{j\cdot}(\hat{t})\hat{x} - b_j(\hat{t}) + \tau B_{j\cdot}(\hat{t})d < 0 \ , \ \ j \in J\setminus J_0 & \forall \text{ small } \tau > 0 \ .
\end{array}$$

Looking at the last relations, we require the feasibility conditions $B_{j\cdot}(\hat{t})x_\tau - b_j(\hat{t}) \leq 0$, $j \in J \setminus J_0$, and define

$$\tau_0 := \min_{\substack{j\in J\setminus J_0 \\ B_{j\cdot}(\hat{t})d>0}} \frac{b_j(\hat{t}) - B_{j\cdot}(\hat{t})\hat{x}}{B_{j\cdot}(\hat{t})d} > 0 \ .$$

Then we obviously obtain (cf., also (6.78)) that the set

$$S(\hat{t}) := \{\hat{x} + \tau d \mid 0 \leq \tau \leq \tau_0\}$$

consists of gc-points of $L(t)$. In case $\tau_0 = \infty$, $S(\hat{t})$ is a half-line. In case $\tau_0 < \infty$, as we shall show, generically the min-value defining τ_0 is attained at precisely one index $j_1 \in J \setminus J_0$. Note that the conditions rank $B_{J_0\setminus\{j_0\}}(\hat{t}) = n-1$, $B_{J_0\setminus\{j_0\}}(\hat{t})d = 0$, and $B_{j_1\cdot}(\hat{t})d > 0$, imply that the matrix $B_{J_0\setminus\{j_0\}\cup\{j_1\}}(\hat{t})$ is nonsingular. To see that the minimizer defining τ_0 is attained only at one index j_1, assume to the contrary that τ_0 is attained at j_1 and j_2, $j_1 \neq j_2$. Then at the gc-point $(\tilde{x}, \hat{t})$ with $\tilde{x} := \hat{x} + \tau_0 d$ the following relations hold (cf., (6.78)):

$$\begin{array}{lll}
B_{J_0\setminus\{j_0\}}(\hat{t})^T\mu & & = -c(\hat{t}) \\
 & B_{J_0\setminus\{j_0\}\cup\{j_1\}}(\hat{t})\tilde{x} & = b_{J_0\setminus\{j_0\}\cup\{j_1\}}(\hat{t}) \\
 & B_{j_2\cdot}(\hat{t})\tilde{x} & = b_{j_2}(\hat{t})
\end{array} \ .$$

However with the same arguments as in the case (5d) for (6.77), these relations yield two linearly independent equations, and can be generically excluded.
If we take the intersection $\mathcal{P}_r$ of all generic subsets $\mathcal{P}_{a2}, \mathcal{P}_{a3}, \mathcal{P}_{a4}, \mathcal{P}_{b1}, \mathcal{P}_{b2}, \mathcal{P}_{51}$ of $\mathcal{P}$ we have found a generic subset $\mathcal{P}_r$ of $\mathcal{P}$ with the required conditions.

We finally note, that the conditions leading to points of Type 5, Type 2^* in (a) and Type 4 at infinity in (b) correspond to the fact that the problem $(c, B, b) \in \mathcal{P}$ transversally meets at $\hat{t}$ a manifold in $\mathbb{R}^n \times M^{m,n} \times \mathbb{R}^m$ of codimension 1. So generically these situations can only occur at a discrete set of point $\hat{t}$ in $\mathbb{R}$.

□

It is clear that a result as in Theorem 6.14 is also valid if we define instead of (6.66):

$$\hat{t} = \inf \{t \in I(\overline{x}, \overline{t})\} \quad \text{and consider} \quad \lim_{t \downarrow \hat{t}} x(t) \ .$$

As for the quadratic program $Q(t)$, also for $L(t)$ generically, a point of Type 4 at infinity is two-sided if $J = J_{(\overline{x},\overline{t})}$ holds.

THEOREM 6.15. [Two-sided point of Type 4 at infinity]
There is a generic subset $\mathcal{P}_r$ of $\mathcal{P}$ such that for any $(c, B, b) \in \mathcal{P}_r$ in addition to the properties in Theorem 6.14 the following holds at any point of Type 4 at infinity as in Theorem 6.14(b), so that the corresponding branch of Type 1 points $(x(t), t),\ t < \hat{t}$, satisfies $J_{(x(t),t)} = J_{(\overline{x},\overline{t})}$, $|J_{(\overline{x},\overline{t})}| = n$, and $\lim_{t\uparrow\hat{t}} \frac{x(t)}{\|x(t)\|} = \tilde{x}$.
in case $J_{(\overline{x},\overline{t})} = J$: there exists a branch of Type 1 points of $L(c, B, b)$, parameterized by a C^∞-curve $(x(t), t)$ for $\hat{t} < t < \hat{t} + \varepsilon$, for some $\varepsilon > 0$, such that $J_{(x(t),t)} = J_{(\overline{x},\overline{t})}$ and

$$\lim_{t\downarrow\hat{t}} \frac{x(t)}{\|x(t)\|} = -\tilde{x} \ .$$

In case $J_{(\overline{x},\overline{t})} \subsetneq J$: this branch right to $\hat{t}$ is generically excluded.

Proof. Recall that we have given a curve of Type 1 points $(x(t), t))$ of $L(c, B, b)$ such that for $t < \hat{t}$ (t close to $\hat{t}$) with $J_0 := J_{(\overline{x},\overline{t})}$, $|J_0| = n$, the matrix $B_{J_0}(t)$ is nonsingular and $x(t)$ is the unique solution of

$$B_{J_0}(t)x(t) = b_{J_0}(t) \ , \qquad t < \hat{t} \ , \tag{6.79}$$

with components given by Cramer's rule (see (7.3)) as,

$$x_i(t) = \frac{\det B^i_{J_0}(b_{J_0})(t)}{\det B_{J_0}(t)} \ , \tag{6.80}$$

where $B^i_{J_0}(b_{J_0})(t)$ is the matrix obtained from $B_{J_0}(t)$ by replacing the i-th column by $b_{J_0}(t)$.
Now, condition (4b) of Type 4 at infinity implies rank $B_{J_0}(\hat{t}) = n - 1$, as well as rank $\big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) = n$. Consequently, there is some i_0 such that the i_0-th column $B_{\cdot i_0}(\hat{t})$ of $B_{J_0}(\hat{t})$ is linearly dependent from the $n-1$ other columns. So for the matrix $B^{i_0}_{J_0}(\hat{t})$ obtained from $B_{J_0}(\hat{t})$ by skipping column $B_{\cdot i_0}(\hat{t})$ it follows with the matrix $B^{i_0}_{J_0}(b_{J_0})(t)$ above:

$$\text{rank } B^{i_0}_{J_0}(b_{J_0})(\hat{t}) = \text{rank } \big(B^{i_0}_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) = \text{rank } \big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) = n \ .$$

Thus $\det B^{i_0}_{J_0}(b_{J_0})(\hat{t}) \neq 0$ is valid. We can assume that i_0 is the index with max value $|\det B^{i_0}_{J_0}(b_{J_0})(\hat{t})| > 0$. In view of $\det B_{J_0}(\hat{t}) = 0$, from (6.80) we conclude $\lim_{t\uparrow\hat{t}} x_{i_0}(t) = \infty$ and with the solution $\tilde{x}$ in (6.73) (proof of Theorem 6.14(b)) we find

$$\lim_{t\uparrow\hat{t}} \frac{x_{i_0}(t)}{\|x(t)\|} = \tilde{x}_{i_0} \neq 0 \ . \tag{6.81}$$

We now show that there is a generic subset $\mathcal{P}_4$ of $\mathcal{P}$ such that for any $(c, B, b) \in \mathcal{P}_4$ at all points $\hat{t}$ of Type 4 at infinity we must have

$$\frac{d}{dt} \det B_{J_0}(\hat{t}) \neq 0 \ . \tag{6.82}$$

This implies rank $B_{J_0}(t) = n$ for all $t \neq \hat{t}$, t close to $\hat{t}$, and (6.79) defines also for $t,\ \hat{t} < t < \hat{t} + \varepsilon$ (some $\varepsilon > 0$) a solution curve $(x(t), t)$.
To prove (6.82) we apply Ex.5.9 with $S = M^n(n-1)$, $f(t) = B_{J_0}(t),\ h(M) = \det M$, $M \in M^n$ $(t = x)$, as well as Theorem 6.1 with $j^0 f =$ graph (f). This assures that there is a generic subset $\mathcal{P}_4$ of $\mathcal{P}$, such that for any $(c, B, b) \in \mathcal{P}_4$ we have $f \pitchfork M^n(n-1)$, or equivalently (cf., Ex 5.9)

$$0 \neq \frac{d}{dt} h(f(\hat{t})) = \frac{d}{dt} \det B_{J_0}(\hat{t}) \ .$$

<u>case $J_0 = J$:</u> In this case also these points $x(t),\ t > \hat{t}$, are feasible for $L(t) = L(c, B, b)$ (vertices) and the points $(x(t), t)$ are of Type 1 with $J_{(x(t),t)} = J_0 = J_{(\overline{x},\overline{t})}$.
With the same arguments as before (6.73), in view of rank $B_{J_0}(\hat{t}) = n-1$, there is (up to a sign) a unique $\hat{x}$, $\|\hat{x}\| = 1$, satisfying

$$B_{J_0}(\hat{t})\hat{x} = 0 \ , \qquad \lim_{t \downarrow \hat{t}} \frac{x(t)}{\|x(t)\|} = \hat{x} \ .$$

Compared with (6.73) we must have $\hat{x} = \pm\tilde{x}$. But since by (6.82) the value $\det B_{J_0}(t)$ changes its sign when passing $\hat{t}$, together with (6.81) we can conclude $\hat{x} = -\tilde{x}$, which proves the statement for $J = J_0$.

<u>case $J_0 \subsetneq J$:</u> Then we can take an index $j_0 \in J \setminus J_0$ (recall $J_0 = J_{(\overline{x},\overline{t})}$). For a generic subset $\mathcal{P}_{42}$ of $\mathcal{P}$ the solution $\tilde{x},\ \|\tilde{x}\| = 1$, of $B_{J_0}(\hat{t})\tilde{x} = 0$ (cf., (6.72) and (6.73)) satisfies

$$B_{j_0\cdot}(\hat{t})\tilde{x} \neq 0 \ . \tag{6.83}$$

Indeed, assuming that $\tilde{x}$ satisfies $B_{J_0}(\hat{t})\tilde{x} = 0$ and $B_{j_0\cdot}(\hat{t})\tilde{x} = 0$, then the $(n+1) \times n$-matrix $\binom{B_{J_0}(\hat{t})}{B_{j_0\cdot}(\hat{t})}$ has rank $n-1$, i,e, (c, B, b) meets at $\hat{t}$ a submanifold of codimension 2, which can generically be excluded.
Now consider the curve $(x(t), t)$, $t \neq \hat{t}$ of solutions of (see (6.79) and (6.82))

$$B_{J_0}(t)x(t) = b_{J_0}(t) \ , \quad t \neq \hat{t} \ . \tag{6.84}$$

As shown above, it holds

$$\lim_{t \uparrow \hat{t}} \frac{x(t)}{\|x(t)\|} = \tilde{x} \ , \quad \lim_{t \downarrow \hat{t}} \frac{x(t)}{\|x(t)\|} = -\tilde{x} \ . \tag{6.85}$$

By (6.83) it follows $|B_{j_0\cdot}(\hat{t})\tilde{x}| =: \delta > 0$. Suppose now, $B_{j_0\cdot}(\hat{t})\tilde{x} = \delta > 0$. Then by continuity, using $\lim_{t \to \hat{t}} \|x(t)\| = \infty$, it follows from (6.85) for $t < \hat{t}$, t close to $\hat{t}$,

$$B_{j_0\cdot}(\hat{t}) \frac{x(t)}{\|x(t)\|} \geq \frac{\delta}{2} \ , \quad \text{and} \quad B_{j_0\cdot}(\hat{t})x(t) \geq \frac{\delta}{2}\|x(t)\| \to \infty \text{ for } t \uparrow \hat{t} \ ,$$

and the points $x(t),\ t < \hat{t}$, t close to $\hat{t}$, would be infeasible, contradicting our assumption that $(x(t), t),\ t < \hat{t}$, are Type 1 points of $L(c, B, b)$ ($\hat{t}$ is a left-sided point of Type 4 at infinity). So we must have $B_{j_0\cdot}(\hat{t})(-\tilde{x}) = \delta > 0$ and with the same reasoning, in view of the second relation in (6.85) for $t > \hat{t}$, t close to $\hat{t}$, we find

$$B_{j_0\cdot}(t)\frac{x(t)}{\|x(t)\|} \geq \frac{\delta}{2}\ , \quad \text{and} \quad B_{j_0\cdot}(t)x(t) \geq \frac{\delta}{2}\|x(t)\| \to \infty \text{ for } t \downarrow \hat{t}\ .$$

But this means that the points $x(t)$, $t > \hat{t}$, of the solution curve $(x(t), t)$ of (6.84) become infeasible for $t \downarrow \hat{t}$. Consequently in this case $J_0 \subsetneq J$ a two-sided point of Type 4 at infinity is generically excluded.

□

REMARK 6.3. In the formulation of our main Theorem 6.14 we tacitly assumed $m \geq n$. In the case $m \leq n - 1$ there does not exist any Type 1 point (primal vertex). Note that for $m < n - 1$ generically there does not occur any gc-point, and for $m = n - 1$, generically, only for a discrete set of points $\hat{t} \in \mathbb{R}$ a line

$$S(\hat{t}) = \{\overline{x} + \tau d \mid \tau \in \mathbb{R}\}$$

of gc-points $x_\tau = \overline{x} + \tau d$, $\tau \in \mathbb{R}$, $x_\tau \in \mathcal{F}(\hat{t})$ may appear (see (6.63)) such that with $J = J_{(x_\tau, \hat{t})}$, $|J| = n - 1$, and some $\mu_j \in \mathbb{R},\ j \in J$, it holds (see Figure 6.53)

$$c(\hat{t}) + \sum_{j \in J} \mu_j B_{j\cdot}(\hat{t})^T = 0\ , \qquad B_J(\hat{t})d = 0\ .$$

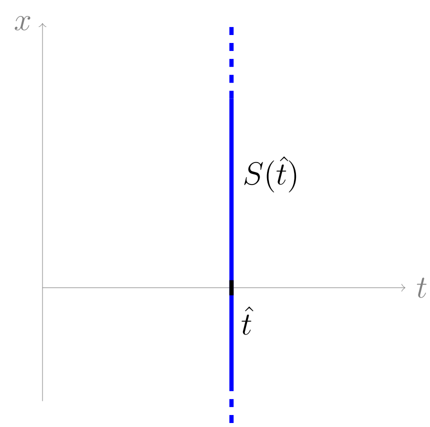

FIGURE 6.53. A line of gc-points $S(\hat{t})$ for $m = n - 1$.

Let us finally comment on the generic behaviour of the set

$$C_m = \{(x, t) \mid x \text{ is a local minimizer of } L(t)\} \subset C_{gc}\ .$$

Clearly if $(\overline{x}, \overline{t}) \in C_m$ is a gc-point of Type 1 of $L(c, B, b)$, then generically there exists a maximal component of points of Type 1 in C_m, parameterized by a curve $(x(t), t)$, $t \in I(\overline{x}, \overline{t})$, where $I(\overline{x}, \overline{t}) \subset \mathbb{R}$ is a maximal open interval.

Around a gc-point of Type 5, generically, the sets C_{gc}, C_m show a structure as for $Q(t)$ or $P(t)$ (cf., Figure 6.24, Figure 6.25, Figure 6.26, Figure 6.27).

Structur of C_m near a point $\hat{t}$ of Type 4 at infinity.

Let us also look at the generic behavior of the set C_m around a point $\hat{t}$ of Type 4 at infinity. So suppose, the curve $(x(t), t),\ t \in I(\overline{x}, \overline{t})$, $t < \hat{t}$, consists of Type 1 points corresponding to (global) minimizers of $L(t) = L(c, B, b)$. Recall that the minimizers $x(t)$ and corresponding multiplier vectors $\mu(t)$ are given as unique solutions of

$$B_{J_0}(t)x(t) = b_{J_0}(t) \tag{6.86}$$

$$B_{J_0}(t)^T\mu(t) = -c(t) \tag{6.87}$$

where $J_0 = J_{(\overline{x},\overline{t})}$, $|J_0| = n$ (see condition (4a) above). If $\hat{t}$ is a one-sided (left-sided) point of Type 4 at infinity, then near $\hat{t}$ the set C_m contains the curve $(x(t), t),\ \ t \in I(\overline{x}, \overline{t})$ as given in Figure 6.54.

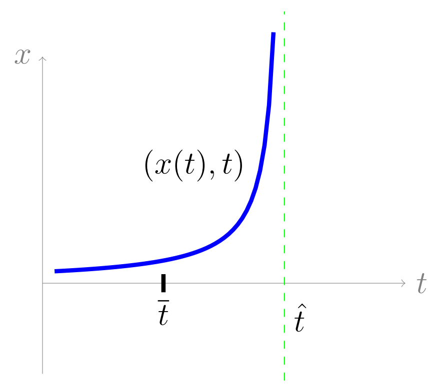

FIGURE 6.54. Part of C_m near a left-sided point $\hat{t}$ of Type 4 at infinity, where $\overline{x}$ is a minimizer of $L(\hat{t})$.

Now suppose that $\hat{t}$ is a two-sided point of Type 4 at infinity. By Theorem 6.15, then we must have $J_0 = J$. Recall that for such a point $\hat{t}$ we have shown (see the proof of Theorem 6.14 part (b)) that generically it holds

$$\text{rank } B_{J_0}(\hat{t}) = n - 1, \qquad \text{rank } \big(B_{J_0}(\hat{t})\ b_{J_0}(\hat{t})\big) = n\ .$$

In the same way we can prove, that generically at $\hat{t}$ the condition (see (6.87))

$$\text{rank } \big(B_{J_0}(\hat{t})^T\ \ -c(t)\big) = n$$

holds. The proof of Theorem 6.15 was based on the analysis of the solution $x(t)$ of (6.86) for $t \neq \hat{t}$. The components $x_i(t)$ of $x(t)$ are given in (6.80). We can apply the same arguments to the solution $\mu(t)$ of (6.87) given explicitly by

$$\mu_i(t) = \frac{\det B^i_{J_0}(t)^T}{\det B_{J_0}(t)^T}\ , \quad t \neq \hat{t},\ t \approx \hat{t}\ ,$$

where $B^i_{J_0}(t)^T$ denotes the matrix obtained from $B_{J_0}(t)^T$ by replacing the i-th column $B_{i\cdot}(t)^T$ by $-c(t)$. We then find the following: $\lim_{t\uparrow\hat{t}} \|\mu(t)\| = \infty$ and for some $\tilde{\mu}$, $\|\tilde{\mu}\| = 1$,

$$\lim_{t\uparrow\hat{t}} \frac{\mu(t)}{\|\mu(t)\|} = \tilde{\mu} \,, \qquad \lim_{t\downarrow\hat{t}} \frac{\mu(t)}{\|\mu(t)\|} = -\tilde{\mu} \,.$$

Since $x(t),\ t < \hat{t}$, are minimizers, the condition $\mu(t) > 0$ must hold for $t < \hat{t}$, implying $\tilde{\mu} \geq 0$. So $-\tilde{\mu}$ must contain a negative component $-\tilde{\mu}_{i_0} < 0$, and therefore $\mu_{i_0}(t) < 0$ must be valid for $t > \hat{t}$, t close to $\hat{t}$. But recalling that the solution $x(t)$ of (6.86) for $t > \hat{t}$ are gc-points of $L(t)$ with corresponding multipliers $\mu(t)$ given by (6.87), we conclude that these points $x(t)$ "right to $\hat{t}$" cannot be minimizers. Consequently the set C_m of minimizers of $L(t)$ only contains the points $(x(t), t),\ t \in I(\overline{x}, \overline{t})$, $t < \hat{t}$, as shown in Figure 6.55.

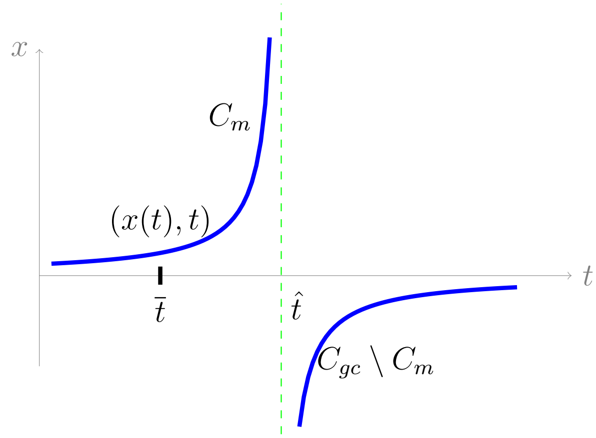

FIGURE 6.55. Part of C_m near a two-sided point $\hat{t}$ of Type 4 at infinity, where $x(\overline{t})$ is a minimizer of $L(\overline{t})$. The points $(x(t), t),\ t > \hat{t}$ belong to $C_{gc} \setminus C_m$.

Structur of C_m near a point $(\hat{x}, \hat{t})$ of Type 2*.

We finally analyse the generic behavior of the set C_m near a point $(\hat{x}, \hat{t})$ of Type 2* (see the definition of this Type, and the proof of Theorem 6.14, case (a4)).

We first consider the case $\tau_0 < \infty$: Here with $\tilde{x} = \hat{x} + \tau_0 d$ (d in (6.65)) the set $S(\hat{t})$ is given by

$$S(\hat{t}) = \{(1-\tau)\hat{x} + \tau\tilde{x} \mid \tau \in [0,1]\} \,.$$

For notational simplicity we put (see condition (2c) of Type 2*)

$$\begin{aligned} J_0 = J_{(\hat{x},\hat{t})} &:= \{1, \ldots, n\},\ j_0 := 1,\ j_1 = n+1 \,, \\ J_1 = J_{(\tilde{x},\hat{t})} &:= \{n+1, 2, \ldots, n\},\ J_2 := \{2, \ldots, n\} \,. \end{aligned}$$

So, we have the following situation for $t \approx \hat{t}$:

(6.88) $$\text{rank } B_{J_0}(t) = \text{rank } B_{J_1}(t) = n \,, \quad \text{rank } B_{J_2}(t) = n - 1 \,.$$

We now consider the unique solutions $x^0(t),\ x^1(t)$ of

$$B_{J_0}(t)x^0(t) = b_{J_0}(t)\ , \qquad B_{J_1}(t)x^1(t) = b_{J_1}(t)\ , \quad t \approx \hat{t}\ . \tag{6.89}$$

By construction, $x^0(t),\ x^1(t),\ t \neq \hat{t}$, are Type 1 vertices of $L(t)$ satisfying $x^0(\hat{t}) = \hat{x}$, $x^1(\hat{t}) = \tilde{x}$, with corresponding Lagrangean multipliers $\mu^0(t),\ \mu^1(t)$ given as unique solutions of

$$B_{J_0}(t)^T\mu^0(t) = -c(t)\ , \qquad B_{J_1}(t)^T\mu^1(t) = -c(t)\ . \tag{6.90}$$

There also is a unique vector $y(t)$ solving

$$B_{J_0}(t)^T y(t) = B_{n+1\cdot}(t)^T\ . \tag{6.91}$$

By Cramer's rule the solutions $\mu^0(t),\ \mu^1(t),\ y(t)$ are given by

$$\mu^0_j(t) = \frac{\det B^j_{J_0}(t)^T}{\det B_{J_0}(t)^T}\ , \quad \mu^1_j(t) = \frac{\det B^j_{J_1}(t)^T}{\det B_{J_1}(t)^T}\ , \quad y_j(t) = \frac{\det \tilde{B}^j_{J_0}(t)^T}{\det B_{J_0}(t)^T}\ , \tag{6.92}$$

where $B^j_{J_0}(t)^T,\ B^j_{J_1}(t)^T,\ \tilde{B}^j_{J_0}(t)^T$, resp., are obtained by replacing the j-th column of $B_{J_0}(t)^T,\ B_{J_1}(t)^T,\ B_{J_0}(t)^T$, resp., by $-c(t),\ -c(t),\ B_{n+1\cdot}(t)^T$, respectively.
Assuming that the vertices $x^0(t),\ t < \hat{t}$, are minimizers of $L(t)$, we must have

$$\mu^0_j(t) > 0,\ j = 2,\dots,n,\ t \approx \hat{t}, \quad \mu^0_1(t) > 0,\ t < \hat{t},\ \mu^0_1(\hat{t}) = 0\ . \tag{6.93}$$

The condition $\mu^0_1(\hat{t}) = 0$ implies $\det B^1_{J_0}(\hat{t})^T = 0$. By definition (see (6.92)) we find

$$\det B^1_{J_0}(t)^T = \det B^1_{J_1}(t)^T \quad \text{and thus} \quad \mu^0_1(\hat{t}) = \mu^1_1(\hat{t}) = 0\ , \tag{6.94}$$

and further using (6.88), (6.91),

$$\begin{aligned} 0 \neq \det B_{J_1}(t)^T &= \det\bigl(B_{n+1\cdot}(t)^T\ B_{2\cdot}(t)^T \dots B_{n\cdot}(t)^T\bigr) \\ &= \det\bigl(B_{J_0}(t)^T y(t)\ B_{2\cdot}(t)^T \dots B_{n\cdot}(t)^T\bigr) && (6.95) \\ &= y_1(t)\det\bigl(B_{1\cdot}(t)^T\ B_{2\cdot}(t)^T \dots B_{n\cdot}(t)^T\bigr) = y_1(t)\det B_{J_0}(t)^T\ , && (6.96) \end{aligned}$$

and thus $y_1(t) \neq 0$ (for $t \approx \hat{t}$) and then (cf., (6.92), (6.94)) $\mu^1_1(t) = \mu^0_1(t)/y_1(t)$. Also for $j \geq 2$ (use (6.92)):

$$\begin{aligned} \det B^j_{J_1}(t)^T &= \det\bigl(B_{n+1\cdot}(t)^T\ B_{2\cdot}(t)^T .. \underbrace{-c(t)}_{j\text{-th}} .. B_{n\cdot}(t)^T\bigr) \\ &= \det\bigl(y_1(t)B_{1\cdot}(t)^T + y_j(t)B_{j\cdot}(t)^T\ B_{2\cdot}(t)^T .. \underbrace{-c(t)}_{j\text{-th}} .. B_{n\cdot}(t)^T\bigr) \\ &= y_1(t)\det B^j_{J_0}(t)^T + y_j(t)\det\bigl(B_{j\cdot}(t)^T\ B_{2\cdot}(t)^T .. \underbrace{-c(t)}_{j\text{-th}} .. B_{n\cdot}(t)^T\bigr)\ . \end{aligned}$$

In view of (6.92), (6.96) this implies for $j \geq 2$,

$$\mu^1_j(t) = \mu^0_j(t) + \frac{y_j(t)}{y_1(t)} \cdot \frac{\det\bigl(B_{j\cdot}(t)^T\ B_{2\cdot}(t)^T .. \overbrace{-c(t)}^{j\text{-th}} .. B_{n\cdot}(t)^T\bigr)}{\det B_{J_0}(t)^T}\ . \tag{6.97}$$

From (6.90) using $\mu_1^0(\hat{t}) = 0$ it follows

$$\det\big(B_{j\cdot}(\hat{t})^T \; B_{2\cdot}(\hat{t})^T .. \underbrace{-c(\hat{t})}_{j\text{-th}} .. B_{n\cdot}(\hat{t})^T\big) = 0 \,,$$

so that from (6.97) by continuity we conclude (see (6.93))

(6.98) $$\mu_j^1(t) > 0, \quad j = 2, \dots, n \quad \text{for } \; t \approx \hat{t} \,.$$

Finally as in the proof of (6.82) we generically can assume

$$\frac{d}{dt} \det B_{J_0}^1(\hat{t})^T \neq 0 \quad \text{and} \quad \frac{d}{dt} \det B_{J_1}^1(\hat{t})^T \neq 0 \,,$$

which means (cf., (6.92)) that the values of $\mu_1^0(t)$, $\mu_1^1(t)$ change sign when passing $\hat{t}$. More precisely,

$$\mu_1^0(t) \;\begin{matrix} > 0 & \text{for } t < \hat{t} \\ < 0 & \text{for } t > \hat{t} \end{matrix} \,, \quad \mu_1^1(t) \;\begin{matrix} < 0 & \text{for } t < \hat{t} \\ > 0 & \text{for } t > \hat{t} \end{matrix} \,.$$

Together with (6.98) this shows:

$$\begin{matrix} \text{for } \; t < \hat{t}: & x^0(t) \text{ are minimizers, } x^1(t) \text{ are not minimizers} \\ \text{for } \; t > \hat{t}: & x^1(t) \text{ are minimizers, } x^0(t) \text{ are not minimizers} \end{matrix} \,,$$

so that near $(\hat{x}, \hat{t})$, $(\tilde{x}, \hat{t})$ the set C_m (in blue) looks like sketched in Figure 6.56.

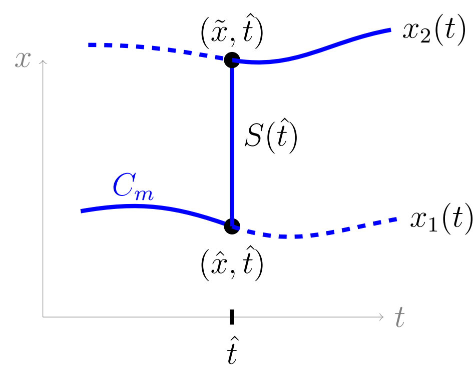

FIGURE 6.56. Set C_m around a Type 2* point $(\hat{x}, \hat{t})$ where $S(\hat{t})$ is a line-segment.

In case $\tau_0 = \infty$, the same holds for the solution $x^0(t)$ of (6.89) with corresponding multiplier $\mu^0(t)$ in (6.90). Here, we also have $\mu_j^0(t) > 0$, $j \geq 2$, for $t \approx \hat{t}$, and $\mu_1^0(t)$ changes sign at $\hat{t}$ from positive to negative. So we obtain

(6.99) $$\mu_1^0(t) < 0 \quad \text{for } \; t > \hat{t} \,.$$

Now we consider the (unique) solution $d(t)$ of

$$B_{J_0}(t) d(t) = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \in \mathbb{R}^n \,, \quad t \approx \hat{t} \,.$$

For $t = \hat{t}$ the vector $d(\hat{t})$ coincides with the direction d in (6.65). Then with the gc-points $x^0(t)$ and the corresponding multipliers $\mu^0(t)$ in (6.90) it follows, using (6.99),

$$c(t)^T d(t) = -\mu^0(t)^T B_{J_0}(t) d(t) = \mu_1^0(t) < 0 \quad \text{for } t > \hat{t}\ .$$

This means that for $t > \hat{t}$, t close to $\hat{t}$, the vector $d(t)$ is a descent direction for $c(t)^T x$ at $x^0(t)$. Consequently, for $t > \hat{t}$, the objective of the program $L(t)$ is unbounded. In particular, near $(\hat{x}, \hat{t})$ the set C_m (in blue) looks as given in Figure 6.57.

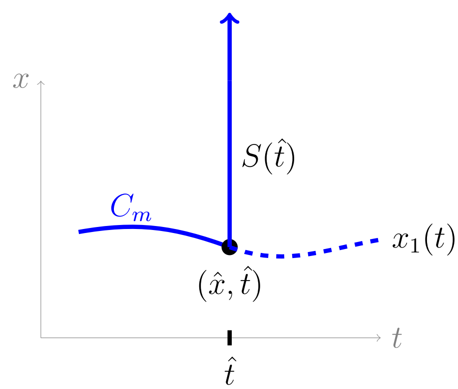

FIGURE 6.57. Set C_m around a Type 2* point $(\hat{x}, \hat{t})$ where $S(\hat{t})$ is a half-line.

CHAPTER 7

Appendix: Basics from linear algebra and analysis

To keep the booklet largely selfcontained, in this appendix we list some facts from linear algebra and calculus which are used regularly.

Determinant, inverse of a matrix, Cramer's rule. For a matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ we denote its determinant by $\det(A)$. Here is the well-known explicit Leibniz formula,

$$\det(A) = \sum_{\sigma \in \pi_n} (\text{sgn}\sigma)\ a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}\ , \tag{7.1}$$

where the sum is taken over all $n!$ permutations σ of $\{1, \ldots, n\}$ and $(\text{sgn}\sigma) = \pm 1$ is the so-called sign of σ.
Other well-known formulas are: $\det(A^T) = \det(A)$ and in case A is nonsingular,

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A) \cdot \det(D - CA^{-1}B)\ . \tag{7.2}$$

The latter formula follows from the relation,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ C & D - CA^{-1}B \end{pmatrix}\ .$$

Also the next relation is standard:

$$\det(A) \neq 0 \quad \Leftrightarrow \quad \text{the inverse } A^{-1} \text{ of } A \text{ exists}\ .$$

For an invertible matrix A we have

$$A^{-1} = \frac{1}{\det(A)} A_{adj} \quad \text{where} \quad A_{adj} = (m_{ij})^T$$

with minors $m_{ij} = (-1)^{i+j} \det(A(i,j))$, and $A(i,j)$ is the $(n-1) \times (n-1)$-matrix obtained from A by deleting row i and colomn j. This formula implies, for invertible A, the following representation of the unique solution x of $Ax = b$:

$$x = A^{-1}b \quad \text{with components} \quad x_i = \frac{\det(A_i(b))}{\det(A)} \qquad \text{Cramer's rule}, \tag{7.3}$$

where $A_i(b)$ is the matrix obtained from A by replacing the i-th column $A_{\cdot i}$ of A by b.

Eigenvalues.

Given $A \in M^n$, a number $\lambda \in \mathbb{C}$ is called an eigenvalue of A if there exists a nonzero vector $x \in \mathbb{C}^n$ such that

$$Ax = \lambda x \ .$$

Such a pair (x, λ) is also called eigenpair of A.
We recall that a real-valued matrix $A \in M^n$ may have complex eigenvalues and eigenvectors. However if A is a symmetric realvalued matrix, i.e., if $A \in S^n$, then all eigenvalues and eigenvectors are real. It is also well-known that for $A \in S^n$ the eigenpairs $(x_i, \lambda_i) \in \mathbb{R}^n \times \mathbb{R}, \ i = 1, \ldots, n$, can be chosen such that the eigenvectors $x_i \in \mathbb{R}^n$ are orthonormal:

$$x_i^T x_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \ .$$

Ex. 7.1. *Let $A \in S^n$. Show that eigenvectors x_1, x_2 corresponding to eigenvalues $\lambda_1 \neq \lambda_2$ are orthogonal.*

Proof. The statement follows from (use $A^T = A$)

$$\begin{aligned} (\lambda_1 - \lambda_2) x_1^T x_2 &= (\lambda_1 x_1)^T x_2 - x_1^T(\lambda_2 x_2) = (Ax_1)^T x_2 - x_1^T (Ax_2) \\ &= x_1^T A^T x_2 - x_1^T A x_2 = 0 \ . \end{aligned}$$

So $\lambda_1 \neq \lambda_2$ implies $x_1^T x_2 = 0$.

□

Mean value formula, Taylor's formula.

We make use of the mean value formula (or first order Taylor formula) in different forms. For $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$, a well-known form is,

$$f(y) - f(x) = \nabla f(x)(y - x) + o(\|y - x\|) \ ,$$

or with some $\tau = \tau(x, y)$, $0 < \tau < 1$,

$$f(y) - f(x) = \nabla f(x + \tau(y - x))(y - x) \ .$$

From the Fundamental Theorem of Calculus we obtain the following integral form of the mean value.

LEMMA 7.1. [Mean value formula]
Let be given $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$, D open, $f \in C^1(D, \mathbb{R}^m)$ and $y, x \in D$ such that the whole line segment $\{x + \tau(y - x) \mid \tau \in [0, 1]\}$ is contained in D. Then we have

$$f(y) - f(x) = \int_0^1 \nabla f(x + \tau(y - x))(y - x) \, d\tau \ ,$$

and thus

$$\|f(y) - f(x)\| \leq \max_{0 \leq \tau \leq 1} \|\nabla f(x + \tau(y - x))\| \, \|y - x\| \ . \tag{7.4}$$

Proof. Let $f(x) = (f_i(x), i = 1, \dots, m)^T$ and define $g_i : [0,1] \to \mathbb{R}$ by

$$g_i(\tau) = f_i(x + \tau(y - x)) \ .$$

Then the Fundamental Theorem of Calculus yields

$$g_i(1) - g_i(0) = \int_0^1 g_i'(\tau)\, d\tau \ ,$$

and then by chain rule,

$$\begin{aligned} f_i(y) - f_i(x) &= g_i(1) - g_i(0) = \int_0^1 g_i'(\tau)\, d\tau \\ &= \int_0^1 \nabla f_i(x + \tau(y - x))(y - x)\, d\tau \ . \end{aligned}$$

By definition of the Jacobian this leads to,

$$f(y) - f(x) = \int_0^1 \nabla f(x + \tau(y - x))(y - x)\, d\tau$$

and

$$\begin{aligned} \|f(y) - f(x)\| &\le \int_0^1 \|\nabla f(x + \tau(y - x))\|\, \|y - x\|\, d\tau \\ &\le \max_{0 \le \tau \le 1} \|\nabla f(x + \tau(y - x))\|\, \|y - x\| \ . \end{aligned}$$

□

We also state the second order Taylor formula for C^2-functions $f : \mathbb{R}^n \to \mathbb{R}$ near a point $\overline{x} \in \mathbb{R}^n$ in the form

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x})(x - \overline{x}) + o(\|x - \overline{x}\|^2) \tag{7.5}$$

or with some $\tau = \tau(x, \overline{x})$, $0 < \tau < 1$,

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x}) + \frac{1}{2}(x - \overline{x})^T \nabla^2 f(\overline{x} + \tau(x - \overline{x}))(x - \overline{x}) \ .$$

Inverse- and Implicit Function Theorem. We give some other basis results from analysis (see, e.g., [34, p.221] for a proof).

THEOREM 7.1. [Inverse Function Theorem]
Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a C^k-function ($k \ge 1$) and let $\overline{x} \in \mathbb{R}^n$ such that the Jacobian $\nabla F(\overline{x})$ is nonsingular. Then there are open neighborhoods U of $\overline{x}$ and $V = F(U)$ of $F(\overline{x})$ such that the restriction $F : U \to V$ has a C^k-inverse $F^{-1} : V \to U$ satisfying

$$\nabla F^{-1}(F(x)) = [\nabla F(x)]^{-1} \qquad \textit{for } \ x \in U \ .$$

This Inverse Function Theorem is directly related to the notion of a (local) diffeomorphism. We recall that a C^1-function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to define a local diffeomorphism at $\overline{x} \in \mathbb{R}^n$ if there exists an open neighborhood U of $\overline{x}$ such that $F : U \to V$ with $V := F(U)$, is bijective (thus the inverse F^{-1} exists) and such that the inverse $F^{-1} : V \to U$ is a C^1-function.

REMARK 7.1. Note that if F defines a local diffeomorphism at $\overline{x}$ then by differentiating the identity $F^{-1}(F(x)) = x,\ x \in U$, we obtain

$$\nabla F^{-1}(F(x)) \cdot \nabla F(x) = I\ , \quad x \in U\ .$$

So in particular, the Jacobian $\nabla F(\overline{x})$ must be nonsingular.
In view of the Inverse Function Theorem we thus have for C^1-functions F:

$$\nabla F(\overline{x}) \text{ is nonsingular} \quad \Rightarrow \quad F \text{ defines a local diffeomorphism at } \overline{x}$$

We also present a version of the Implicit Function Theorem. Let be given a system of n equations in $n + p$ variables:

$$F(x,t) = 0\ , \qquad \text{where} \quad F : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n,\ \ (x,t) \in \mathbb{R}^n \times \mathbb{R}^p\ .$$

The Implicit Function Theorem (IFT) makes a statement on the structure of the solution set in the “regular” situation.

THEOREM 7.2. [Implicit Function Theorem (IFT)]
Let $F : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ be a C^1-function $F(x,t)$ with $(x,t) \in \mathbb{R}^n \times \mathbb{R}^p$. Suppose for $(\overline{x}, \overline{t}) \in \mathbb{R}^n \times \mathbb{R}^p$ we have $F(\overline{x}, \overline{t}) = 0$ and the matrix $\nabla_x F(\overline{x}, \overline{t})$ is nonsingular. Then in neighborhoods $U_{\overline{x}}$ of $\overline{x}$, $U_{\overline{t}}$ of $\overline{t}$, the solution set $S(F) := \{(x,t) \mid F(x,t) = 0\}$ is described by a C^1-function $x : U_{\overline{t}} \to \mathbb{R}^n$ such that $x(\overline{t}) = \overline{x}$ and $F(x(t),t) = 0$ for all $t \in U_{\overline{t}}$. More precisely,

$$S(F) \cap U_{\overline{x}} \times U_{\overline{t}} = \{(x(t),t) \mid t \in U_{\overline{t}}\}\ .$$

So, locally near $(\overline{x}, \overline{t})$, the set $S(F)$ is a p dimensional C^1-manifold. Moreover, the gradients $\nabla x(t)$ are given by

$$\nabla x(t) = -[\nabla_x F(x(t),t)]^{-1} \nabla_t F(x(t),t) \quad \textit{for } t \in U_t(\overline{t})\ . \tag{7.6}$$

Proof. For a proof see, *e.g.*, [34]. Note that if $x(t)$ is a C^1-function satisfying $F(x(t),t) = 0$ for $t \approx \overline{t}$, then by differentiation wrt. t we find

$$\nabla_x F(x(t),t) \nabla x(t) + \nabla_t F(x(t),t) = 0\ , \quad t \approx \overline{t}\ ,$$

which yields (7.6). □

EX. 7.2. *Show that in a neighborhood of a solution $(\overline{x}, \overline{t})$ of the system $F(x,t) = 0$ satisfying the assumptions of Theorem 7.2, the solution set $S(F)$ is a manifold of codimension n, and of dimension p in $\mathbb{R}^n \times \mathbb{R}^p$.*

Bibliography

[1] Ahmad F., Still G., Maximization of homogeneous polynomials over the simplex and the sphere: structure, stability, and generic behavior. J. Optim. Theory Appl. 181, no. 3, 972-996, (2019).

[2] Bando S., Urakawa H., Generic properties of the eigenvalue of the Laplacian for compact Riemannian manifolds, Tohoku Math. Journal 35, 155-172, (1983).

[3] Benedetti R., Risler J-J., Real algebraic and semi-algebraic sets, Hermann Éditeurs des sciences et des Arts, Paris, (1990).

[4] Bertsimas D., Tsitsiklis J.N., Introduction to Linear Optimization, Athena Scientific Books, Belmont, (1997).

[5] Bomze I.M., Dickinson P.J.C., Still G., The structure of completely positive matrices according to their CP-rank and CP-plus-rank. Linear Algebra Appl. 482, 191-206, (2015).

[6] Bouza Allende G., Mathematical programs with equilibrium constraints: Solution techniques from parametric optimization. Thesis (Dr.) Universiteit Twente (The Netherlands), (2006).

[7] Bouza G., Guddat J., Still G., Critical sets in one-parametric mathematical programs with complementarity constraints. Optimization 57, no. 2, 319-336, (2008).

[8] Bouza G., Still, G., Solving bilevel programs with the KKT-approach. Math. Program. 138, no. 1-2, Ser. A, 309-332, (2013).

[9] Caron R., Traynor T., The zero set of a polynomial. (2005), www.uwindsor.ca/math/sites/uwindsor.../05-03.pdf

[10] Coste M., An introduction to semi-algebraic geometry. Report, Institut de Recherche Mathématique, Rennes, (2002), https://gcomte.perso.math.cnrs.fr.costelintrotosemialgeo.pdf

[11] Darinka D., On the singularities in linear one-parametric optimization problems. Optimization 22, no. 2, 193-219, (1991).

[12] Dür M., Jargalsaikhan B., Still G., Genericity results in linear conic programming - a tour d'horizon. Math. Oper. Res. 42, no. 1, 77-94, (2017).

[13] Faigle U., Kern W., Still G., Algorithmic Principles of Mathematical Programming, Kluwer, Dordrecht, (2002).

[14] Gibson C.G., Wirthműller K., Du Plessis A.A., Looijenga E.J.N., Topological stability of smooth mappings. Lecture Notes in Math., vol. 552, Springer-Verlag, Berlin (1976).

[15] Günzel H., Jongen H.Th., Stein O., On the closure of the feasible set in generalized semi-infinite programming. CEJOR Cent. Eur. J. Oper. Res. 15, no. 3, 271-280, (2007).

[16] Guillemin V., Pollack A., Differential Topology. Prentice-Hall, Inc., Englewood Cliffs, (1974).

[17] Guddat J., Guerra Vazquez F., Jongen H.Th., Parametric optimization: singularities, pathfollowing and jumps. B. G. Teubner, Stuttgart; John Wiley, Sons, Ltd., Chichester, (1990).

[18] Hirsch, M.W., Differential Topology. Springer, Berlin, (1976).

[19] Jongen H.Th., Jonker P., Twilt, F., Nonlinear optimization in $\mathbb{R}^n$. I. Morse theory, Chebyshev approximation. Methoden und Verfahren der Mathematischen Physik [Methods and Procedures in Mathematical Physics], 29. Verlag Peter D. Lang, Frankfurt am Main, (1983).

[20] Jongen H.Th., Jonker P., Twilt F., Nonlinear optimization in $\mathbb{R}^n$. II. Transversality, flows, parametric aspects. Methoden und Verfahren der Mathematischen Physik [Methods and Procedures in Mathematical Physics], 32. Verlag Peter D. Lang, Frankfurt am Main, (1986).

[21] Jongen H.Th., Jonker P., Twilt F., Nonlinear Optimization in finite Dimensions. Kluwer, Dordrecht, (2000).

[22] Jongen H.Th., Jonker P., Twilt F., Critical sets in parametric optimization. Mathematical Programming 34, p.333-353, (1986).

[23] Jongen H.Th., Jonker P., Twilt F., One-parameter families of optimization problems: equality constraints. Journal of Optimization Theory and Applications 45, No. 1, p.141-160, (1986).

[24] Jongen H.Th., Stein O., On generic one-parametric semi-infinite optimization. SIAM J. Optim. 7, no. 4, 1103-1137, (1997).

[25] Jonker P., Still G., Twilt F., On the partition of real symmetric matrices according to the multiplicities of their eigenvalues. Control and Cybernetics, vol. 23, No. 1/2, 169-181, (1994).

[26] Jonker P., Pouw M., Still G., Twilt F., On the partition of real skew-symmetric matrices according to the multiplicities of their eigenvalues. Charlemagne and his heritage. 1200 years of civilization and science in Europe, Vol. 2 (Aachen, 1995), 439-454, Brepols, Turnhout, (1998).

[27] Jonker P., Still G., Twilt, F. One-parametric linear-quadratic optimization problems. in: Optimization with data perturbations, II. Ann. Oper. Res. 101, 221-253, (2001).

[28] Kato T., Perturbation theory for linear operators, Vol. I and Vol. II. Springer Verlag, Tokyo, (1966).

[29] Lancaster P., Tismenetsky M., The theory of matrices. Academic Press, INC. Boston, (1985).

[30] Lee J.M., Introduction into smooth manifolds. Springer, (2003).

[31] Luenberger, D.G., Linear and Nonlinear Programming. Springer, 2d ed., New York, (2003).

[32] Minerbe V., An introduction to differential geometry, UPCM-Université Paris 6, (2015). https://webusers.imj-prg.fr

[33] Müger M., An introduction to Differential Topology, de Rham Theory and Morse Theory. University of Groningen, www.math.ru.nl/ mueger/diff_notes.pdf

[34] Rudin W., Principles of mathematical analysis (3d ed.). McGraw-Hill, (1976).

[35] Rupp T., Kuhn-Tucker curves for one-parametric semi-infinite programming. Optimization 20, no. 1, 61-77, (1989).

[36] Still G., How to split the eigenvalues of a one-parametric family of matrices. Optimization vol. 49, p. 387-403, (2001).

[37] Still G., Linear bilevel problems: genericity results and an efficient method for computing local minima. Math. Methods Oper. Res. 55, no. 3, 383-400, (2002).

[38] Scholtes S., Stöhr M., How stringent is the linear independence assumption for mathematical programs with complementarity constraints? Math. Oper. Res. 26, no. 4, 851-863, (2001).

[39] Tu L.W., An introduction into manifolds (second edition). Springer, (2010).

[40] Zeidler E., Nonlinear Functional Analysis and Applications IV. Springer (1988).

[41] Zwier G., Structural analysis in semi-infinite programming. Thesis (Dr.) Universiteit Twente (The Netherlands), 165 pp, (1987).

Index