CHAPTER 4

Parametric optimization

In this chapter we start with unconstrained and constrained parametric programs. We analyse the dependence of minimizers on the parameter under assumptions such that the Implicit Function Theorem (IFT) can be applied.
The next section gives a short introduction into parametric equations and states the Implicit Function Theorem which can be seen as the most important basic theorem in parametric optimization.

4.1. Parametric equations and the Implicit Function Theorem

We begin with an illustrative example and consider the non-parametric equation: Solve for $x \in \mathbb{R}$

$$H(x) := x^2 - 2x - 1 = 0 \quad \text{or equivalently} \quad (x-1)^2 - 2 = 0$$

with solutions $x_{1,2} = 1 \pm \sqrt{2}$.
The parametric version is: For parameter $t \in \mathbb{R}$ find a solution $x = x(t)$ of

$$\text{(4.1)} \quad H(x,t) := x^2 - 2t^2 x - t^4 = 0 \quad \text{or equivalently} \quad (x-t^2)^2 - 2t^4 = 0 \ .$$

The solutions are (see Figure 4.1)

$$x_{1,2}(t) = t^2 \pm \sqrt{2}t^2 = t^2(1 \pm \sqrt{2})$$

Obviously the solution curves bifurcate at $(\overline{x}, \overline{t}) = (0,0)$. At this point we find for the partial derivative of H

$$\nabla_x H(\overline{x}, \overline{t}) = 2x - 2t^2_{|\overline{x},\overline{t}} = 0 \ .$$

Rule. For $H : \mathbb{R}^2 \to \mathbb{R}$, $H \in C^1$, the solution set of a one-parametric equation $H(x,t) = 0$ is "normally" locally given by a *one-dimensional* C^1 solution curve $(x(t), t)$. However at points $(\overline{x}, \overline{t})$ where $\nabla_x H(\overline{x}, \overline{t}) = 0$ holds, the solution set may show a *singularity* (such as a bifurcation or nonsmoothness).

The Implicit Function Theorem

More generally we consider systems of n equations in $n + p$ variables:

$$H(x,t) = 0 \qquad \text{where} \quad H : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n, \ \ (x,t) \in \mathbb{R}^n \times \mathbb{R}^p \ .$$

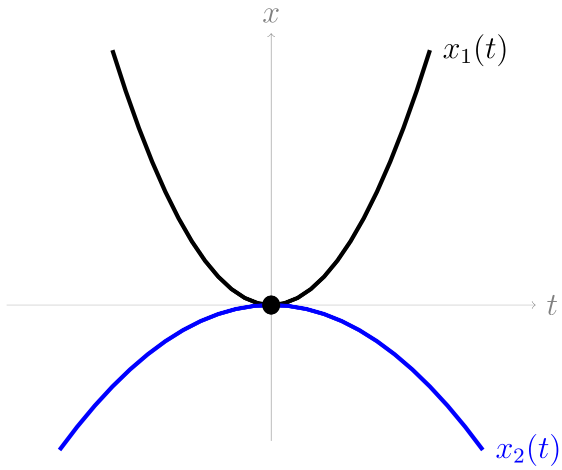

FIGURE 4.1. Sketch of the solution curves $x_1(t), x_2(t)$

The Implicit Function Theorem (IFT) makes a statement on the structure of the solution set of this equation in the "normal" situation.

THEOREM 4.1. *[General version of the IFT]*
Let $H : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ be a C^1-function $H(x,t)$. Suppose for $(\overline{x}, \overline{t}) \in \mathbb{R}^n \times \mathbb{R}^p$ we have $H(\overline{x}, \overline{t}) = 0$ and the matrix

$$\nabla_x H(\overline{x}, \overline{t}) \qquad \textit{is nonsingular.}$$

Then there is a neighborhood $B_\varepsilon(\overline{t})$, $\varepsilon > 0$, of $\overline{t}$ and a C^1-function $x : B_\varepsilon(\overline{t}) \to \mathbb{R}^n$ satisfying $x(\overline{t}) = \overline{x}$ such that near $(\overline{x}, \overline{t})$ the solution set $S(H) := \{(x,t) \mid H(x,t) = 0\}$ is described by

$$\{(x(t), t) \in \mathbb{R}^n \times \mathbb{R}^p \mid t \in B_\varepsilon(\overline{t})\}, \quad \textit{i.e.,} \quad H(x(t), t) = 0 \textit{ for } t \in B_\varepsilon(\overline{t}) \ . \tag{4.2}$$

So, locally near $(\overline{x}, \overline{t})$, the set $S(H)$ is a p dimensional C^1-manifold. Moreover, the gradient $\nabla x(t)$ is given by

$$\nabla x(t) = -[\nabla_x H(x(t), t)]^{-1} \nabla_t H(x(t), t) \quad \textit{for } t \in B_\varepsilon(\overline{t}) \ .$$

Proof. See *e.g.*, [29]. Note that if $x(t)$ is a C^1-function satisfying $H(x(t), t) = 0$, then differentiation wrt. t yields by applying the chain rule,

$$\nabla_x H(x(t), t) \nabla x(t) + \nabla_t H(x(t), t) = 0 \ .$$

□

Pathfollowing in practice

We shortly discuss how a solution curve $(x(t), t)$ of a one-parametric equation

$$H(x,t) = 0 \qquad H : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n, \quad t \in \mathbb{R},$$

can be followed numerically.

The basic idea is to use some sort of Newton procedure. Recall that the classical Newton method is the most fundamental approach for solving a system of n equations in n unknowns:

$$H(x) = 0 \qquad H : \mathbb{R}^n \to \mathbb{R}^n \ .$$

The famous Newton iteration for computing a solution is to start with some (appropriate) starting point x^0 and to iterate according to

$$x^{k+1} = x^k - [\nabla H(x^k)]^{-1} H(x^k), \quad k = 0, 1, \dots \ .$$

It is well-known (see, e.g., [9, Th.11.4]) that this iteration converges quadratically to a solution $\overline{x}$ of $H(\overline{x}) = 0$ if

- x^0 is chosen close enough to $\overline{x}$ and if
- $\nabla H(\overline{x})$ is a nonsingular matrix.

The simplest way to follow approximately a solution curve $x(t)$ of $H(x, t) = 0$, ***i.e.***, $H(x(t), t) = 0$, on an interval $t \in [a, b]$ is to discretize $[a, b]$ by

$$t_\ell = a + \ell \frac{b-a}{N}, \quad \ell = 0, \dots, N \ ,$$

(for some $N \in \mathbb{N}$) and to compute for any $\ell = 0, \dots, N$, a solution $x_\ell = x(t_\ell)$ of $H(x, t_\ell) = 0$ by a Newton iteration,

$$x_\ell^{k+1} = x_\ell^k - [\nabla_x H(x_\ell^k, t_\ell)]^{-1} H(x_\ell^k, t_\ell), \quad k = 0, 1, \dots \ ,$$

starting with $x_\ell^0 = x_{\ell-1} + \frac{(b-a)}{N} x'(t_{\ell-1})$ (for $\ell \geq 1$). The derivative $x'(t_{\ell-1})$ can be computed with the formula in Theorem 4.1. We refer the reader to the book [1] for details *e.g.*, on:

- How to perform pathfollowing efficiently?
- How to deal with branching points $(\overline{x}, \overline{t})$ where different solution curves intersect?

4.2. Parametric unconstrained minimization problems

Let $f(x, t)$ be a C^2-function, $f : \mathbb{R}^n \times T \to \mathbb{R}$, where the parameter set $T \subset \mathbb{R}^p$ is open. We consider the *parametric problem*: for parameter $t \in T$ find a local or global minimizer $x = x(t)$ of

$$P(t): \qquad \min_{x \in \mathbb{R}^n} f(x, t) \tag{4.3}$$

We recall that if for any $t \in T$ the function $f(x, t)$ is convex in x then a (local) minimizer $x(t)$ is a global minimizer of $P(t)$ (cf., Lemma 2.3). In this case we can define the (global) value function of $P(t)$ by

$$v(t) = \inf_{x \in \mathbb{R}^n} f(x, t) \ .$$

In the general nonlinear case, $x(t)$ is thought to be a local minimizer of $P(t)$ and the value function is defined locally by

$$v(t) = f(x(t), t) \ .$$

To solve this problem $P(t)$, for $t \in T$, we have to find solutions x of the *critical point equation*

$$H(x,t) := \nabla_x f(x,t) = 0 \ . \tag{4.4}$$

The next examples show the possible bad behavior in case the regularity condition, that $\nabla_x H(\overline{x}, \overline{t}) = \nabla_x^2 f(\overline{x}, \overline{t})$ is nonsingular, does not hold.

Consider e.g. the parametric program

$$P(t) \quad \min_{x \in \mathbb{R}} \ f(x,t) = tx^2, \qquad \text{for} \qquad t \in \mathbb{R} \ ,$$

near $\overline{t} = 0$. Here for the critical point equation $H(x,t) = 2tx = 0$ we find at $\overline{t} = 0$, $\nabla_x H(x, \overline{t}) = 2\overline{t} = 0$. So the regularity condition fails at all solutions $x \in \mathbb{R}$ of $H(x, \overline{t}) = 0$. The set $S(t)$, of global minimizers is given by:

$$S(t) = \begin{cases} \{0\} & \text{for } t > 0 \\ \mathbb{R} & \text{for } t = 0 \\ \emptyset & \text{for } t < 0 \end{cases} \ .$$

Ex. 4.1. For

$$P(t): \qquad \min_x \ f(x,t) := \frac{1}{3}x^3 - t^2 x$$

the minimizers are given by $x(t) = |t|$ with minimal value $v(t) := f(x(t), t) = -\frac{2}{3}|t|^3$.

Ex. 4.2. *Show that for the parametric minimization problems*

$$P_1(t): \qquad \min_x \ f_1(x,t) := \frac{1}{3}tx^3 - tx$$

$$P_2(t): \qquad \min_x \ f_2(x,t) := \frac{1}{3}t^3x^3 - tx$$

we obtain for the minimizers $x(t)$ and the value function $v(t)$:

for $P_1(t)$: $\qquad x(t) = \begin{cases} -1 & t < 0 \\ \mathbb{R} & t = 0 \\ 1 & t > 0 \end{cases} \ , \ v(t) = \begin{cases} \frac{2}{3}t^3 & t < 0 \\ -\frac{2}{3}t^3 & t > 0 \end{cases}$

for $P_2(t)$: $\qquad x(t) = \begin{cases} -\frac{1}{|t|} & t < 0 \\ \mathbb{R} & t = 0 \\ \frac{1}{|t|} & t > 0 \end{cases} \ , \ v(t) = \begin{cases} \frac{2}{3}t & t < 0 \\ -\frac{2}{3}t & t > 0 \end{cases}$

Note that in both cases the set $S(0)$ of minimizers of $P_1(0), P_2(0)$ is $S(0) = \mathbb{R}$ (see Figure 4.2)

The next theorem describes the solution set of (4.4) near a (non-singular) minimizer $(\overline{x}, \overline{t})$ of (4.3), where the IFT can be applied.

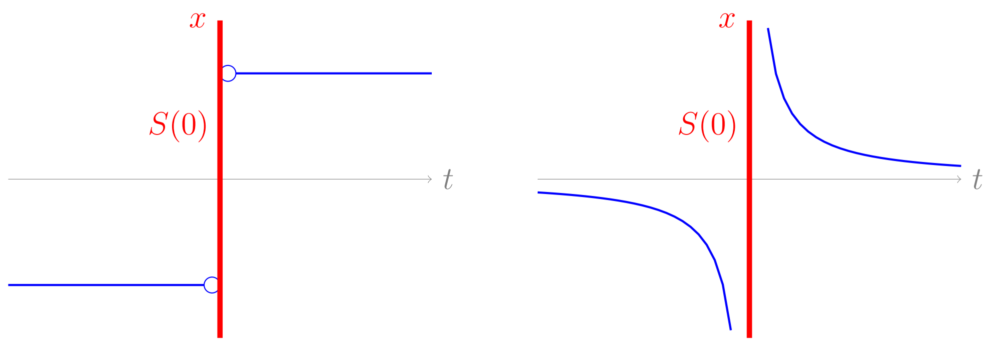

FIGURE 4.2. Minimizers of $P_1(t)$ and minimizers of $P_2(t)$

THEOREM 4.2. *[local stability result based on IFT]*
Let $f(x,t)$ be a C^2-function. Suppose, $\overline{x}$ is a (local) minimizer of $P(\overline{t})$, $\overline{t} \in T$, such that

$$\nabla_x f(\overline{x},\overline{t}) = 0 \quad \textit{and} \quad \nabla_x^2 f(\overline{x},\overline{t}) \succ 0 \quad \textit{(positive definite).}$$

(According to Theorem 2.1, Lemma 2.1, $\overline{x}$ is an isolated strict local minimizer of order 2.)
Then there exists a neighborhood $B_\varepsilon(\overline{t})$, $\varepsilon > 0$, of $\overline{t}$ and a C^1-function $x : B_\varepsilon(\overline{t}) \to \mathbb{R}^n$ such that $x(\overline{t}) = \overline{x}$ and for any $t \in B_\varepsilon(\overline{t})$, $x(t)$ is an (isolated) strict local minimizer of $P(t)$. Moreover for $t \in B_\varepsilon(\overline{t})$,

$$\nabla x(t) = -[\nabla_x^2 f(x(t),t)]^{-1} \nabla_{xt}^2 f(x(t),t) \ ,$$

and the value function $v(t) := f(x(t),t)$ is a C^2-function with

$$\nabla v(t) = \nabla_t f(x(t),t) \quad \textit{and} \quad \nabla^2 v(t) = \nabla_{tx}^2 f(x(t),t) \nabla x(t) + \nabla_t^2 f(x(t),t) \ .$$

Proof. Apply the IFT to the critical point equation $\nabla_x f(x,t) = 0$. □

Remark. Note that for $f \in C^2$ the solution function $x(t)$ is C^1 but the value function $v(t)$ is C^2.

Ex. 4.3. For

$$P(t): \qquad \min\ f(x,t) := \frac{1}{3}x^3 - t^2 x^2 - t^4 x$$

the critical points are given by the curves $x_{1,2}(t) = t^2(1 \pm \sqrt{2})$ and the minimizer by $x_1(t)$ (maximizer by $x_2(t)$) (cf., Figure 4.1). At all points of the solution curves $x_1(t), x_2(t)$ the IFT can be applied except at the branching point at $\overline{t} = 0$.

<u>Rule.</u> The following appears:

- The value function $v(t) = f(x(t),t)$ behaves "smoother" than the minimizer function $x(t)$.

- A singular behavior may appear at solution points $(\overline{x}, \overline{t})$ of $\nabla_x f(x,t) = 0$ where the matrix $\nabla_x^2 f(\overline{x}, \overline{t})$ is singular.

Ex. 4.4. Check the singular behavior for the examples Ex. 4.1 and Ex. 4.2.

4.3. Parametric linear programs

In a parametric linear program (LP) we have given a C^2-matrix function $A(t) : \mathbb{R}^p \to \mathbb{R}^{m\times n}$ with m rows $a_j^T(t), j \in J := \{1,\dots,m\}$, C^2-vector functions $b(t) : \mathbb{R}^p \to \mathbb{R}^m$, $c(t) : \mathbb{R}^p \to \mathbb{R}^n$ and an open parameter set $T \subset \mathbb{R}^p$: For any $t \in T$ we wish to solve the primal program

$$(4.5) \qquad P(t): \quad \max\, c(t)^T x \quad \text{s.t.} \quad x \in F_P(t) = \{x \mid A(t)x \le b(t)\}\ .$$

The corresponding dual reads

$$(4.6) \quad D(t): \quad \min\, b(t)^T y \quad \text{s.t.} \quad y \in F_D(t) = \{y \mid A(t)^T y = c(t), y \ge 0\}\ .$$

For $\overline{t} \in T$ and $\overline{x} \in F_P(\overline{t})$ the active index set is $J_0(\overline{x}, \overline{t}) = \{j \in J \mid a_j(\overline{t})^T \overline{x} = b_j(\overline{t})\}$.

Suppose for $\overline{t} \in T$ the point $\overline{x} \in F_P(\overline{t})$ is a vertex solution of $P(\overline{t})$. To find for t near $\overline{t}$ solutions $x(t)$ of $P(t)$ we have to find feasible solutions x and y of the system of optimality conditions,

$$(4.7) \qquad \text{for } P(t): \quad A_{J_0(\overline{x},\overline{t})}(t)x = b_{J_0(\overline{x},\overline{t})}(t)$$

$$(4.8) \qquad \text{for } D(t): \quad A_{J_0(\overline{x},\overline{t})}(t)^T y_{J_0(\overline{x},\overline{t})} = c(t), \quad y_{J_0(\overline{x},\overline{t})} \ge 0$$

If $\overline{x}$ is a nondegenerate vertex of $P(\overline{t})$, *i.e.*, if $A_{J_0(\overline{x},\overline{t})}(\overline{t})$ is nonsingular, and the minimizer $\overline{y}$ of $D(\overline{t})$ satisfies the strict complementarity (SC) condition $\overline{y}_j > 0, j \in J_0(\overline{x}, \overline{t})$, this is possible by applying the IFT to these systems.

Theorem 4.3. *[Local stability result]*
Let $\overline{x} \in F_P(\overline{t})$ be a vertex maximizer of $P(\overline{t})$ with corresponding dual solution $\overline{y}_{J_0(\overline{x},\overline{t})}$ such that

(1) *$\overline{x}$ is a nondegenerate vertex,* i.e., *LICQ holds,*
(2) $\overline{y}_j > 0,\ \forall j \in J_0(\overline{x}, \overline{t})$. *(SC)*

(According to Theorem 2.4(a), $\overline{x}$ is a maximizer of $P(\overline{t})$ of order $s = 1$.) Then there exist a neighborhood $B_\varepsilon(\overline{t})$, $\varepsilon > 0$, and C^1-functions $x : B_\varepsilon(\overline{t}) \to \mathbb{R}^n$, $y_{J_0(\overline{x},\overline{t})} : B_\varepsilon(\overline{t}) \to \mathbb{R}^{|J_0(\overline{x},\overline{t})|}$ such that $x(\overline{t}) = \overline{x}$, $y_{J_0(\overline{x},\overline{t})}(\overline{t}) = \overline{y}_{J_0(\overline{x},\overline{t})}$ and for any $t \in B_\varepsilon(\overline{t})$ the point $x(t)$ is a vertex maximizer of $P(t)$ (of order s=1) with corresponding multiplier $y_{J_0(\overline{x},\overline{t})}(t)$. Moreover, for $t \in B_\varepsilon(\overline{t})$ the derivatives of $x(t)$ and the value function $v(t) = c(t)^T x(t)$ are given by

$$\nabla x(t) = [A_{J_0(\overline{x},\overline{t})}(t)]^{-1}\left(\nabla b_{J_0(\overline{x},\overline{t})}(t) - \nabla A_{J_0(\overline{x},\overline{t})}(t)x(t)\right)$$

and

$$\nabla v(t) = \nabla c(t)^T x(t) + [y_{J_0(\overline{x},\overline{t})}(t)]^T[\nabla b_{J_0(\overline{x},\overline{t})}(t) - \nabla A_{J_0(\overline{x},\overline{t})}(t)x(t)]\ .$$

Proof. Since the vertex $\overline{x}$ is nondegenerate, the matrix $A_{J_0(\overline{x},\overline{t})}(\overline{t})$ is nonsingular (see Definition 2.3). So the IFT can be applied to the systems (4.7), (4.8), yielding in a neighborhood $B_\varepsilon(\overline{t})$ vertex solutions $x(t)$ of $P(t)$ and $y_{J_0(\overline{x},\overline{t})}(t) > 0$ (since $\overline{y}_{J_0(\overline{x},\overline{t})} = y_{J_0(\overline{x},\overline{t})}(\overline{t}) > 0$) of $D(t)$. Differentiating the relation

$$A_{J_0(\overline{x},\overline{t})}(t)x(t) = b_{J_0(\overline{x},\overline{t})}(t)$$

wrt. t directly leads to the formulas for $\nabla x(t)$ and $\nabla v(t)$. $\square$

Linear production model and shadow prices

We discuss a production model to present a simple application of the result in Theorem 4.3. Assume a factory produces n different products $P_1, \ldots, P_n$. The production relies on material coming from m different resources $R_1, \ldots, R_m$ in such a way that the production of 1 unit of a product P_i requires a_{ji} units of resource R_j.

Suppose we can sell our production for the price of c_i per 1 unit of P_i and that b_j units of each resource R_j are available for the total production. How many units x_i of each product P_i should we produce in order to maximize the total receipt from the sales?

An optimal production plan $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)^T$ corresponds to an optimal solution of the linear program

$$P: \quad \max\ c^T x \quad \text{s.t.} \quad Ax \leq b, \ \ x \geq 0\ . \tag{4.9}$$

Here $A = (a_{ji})$ is the matrix with the elements a_{ji}. Let $\overline{x}$ be a maximizer with corresponding minimizer $\overline{y}$ of the dual problem

$$D: \quad \min\ b^T y \quad \text{s.t.} \quad A^T y \geq c, \ \ y \geq 0\ . \tag{4.10}$$

and maximum profit $\bar{z} \ = \ c^T\overline{x} = b^T\overline{y}$.
Could we possibly increase the profit by spending money on increasing the resource capacity b and adjusting the production plan? If so, how much would we be willing to pay for 1 more unit of resource R_j?
Let us increase for fixed $j_0 \in J_0(\overline{x})$ only the capacity of R_{j_0} as follows. For fixed b we consider the vector $b(t)$ given componentwise by

$$b_{j_0}(t) = b_{j_0} + t, \qquad b_j(t) = b_j,\ j \neq j_0\ , \tag{4.11}$$

depending on the parameter $t \in \mathbb{R}$ near $\overline{t} = 0$. So we have given the parametric linear programs

$$P(t): \ \ \max\ c^T x \quad \text{s.t.} \quad \begin{array}{c} Ax \leq b(t) \\ x \geq 0 \end{array}\ , \ D(t): \ \ \min\ b(t)^T y \quad \text{s.t.} \quad \begin{array}{r} A^T y \geq c \\ y \geq 0 \end{array}$$

for t near $\overline{t} = 0$. We assume that the optimal solutions $\overline{x}$ of $P(\overline{t})$ and $\overline{y}$ of $D(\overline{t})$ satisfy the assumptions in Theorem 4.3. Then, since c, A do not depend on t and (see (4.11))

$$\frac{d}{dt}b_{j_0}(\bar{t}) = 1 \quad \text{and} \quad \frac{d}{dt}b_j(\bar{t}) = 0,\ j \neq j_0 \ ,$$

for the derivative of the value function $v(t)$ in Theorem 4.3 we find

$$\frac{d}{dt}v(\bar{t}) = \sum_{j\in J_0(\overline{x})} \overline{y}_j \frac{d}{dt}b_j(\bar{t}) = \overline{y}_{j_0} \ .$$

This means that near $\bar{t} = 0$ the value function $v(t)$ has a first order approximation

$$v(t) = v(\bar{t}) + \frac{d}{dt}v(\bar{t}) + o(t) = b^T\overline{y} + \overline{y}_{j_0}t + o(t)$$

and thus (note that by assumption $\overline{y}_{j_0} > 0$) we would not want to pay more than $t \cdot \bar{y}_{j_0}$ for t more units of R_{j_0}. In this sense, the coefficients $\bar{y}_j$ of the dual optimal solution $\bar{y}$ can be interpreted as the *shadow prices* of the resources R_j.

REMARK. The notion of shadow prices furnishes also an intuitive interpretation of complementary slackness. If the slack $\bar{s}_i = b_i - \sum_{j=1}^n a_{ij}\bar{x}_j$ is strictly positive at the optimal production $\bar{x}$ (and thus $\overline{y}_i = 0$), we do not use resource R_i to its full capacity. Therefore, we would expect no gain from an increase of R_i's capacity.

4.4. Parametric nonlinear constrained programs

Let $T \subset \mathbb{R}^p$ be some open parameter set. We consider *nonlinear parametric programs* of the form (omitting equality constraints): For $t \in T$ find local minimizers $x = x(t)$ of

$$\text{(4.12)} \quad P(t): \quad \min_x f(x,t) \ \text{ s.t. } \ x \in F(t) = \{x \in \mathbb{R}^n \mid g_j(x,t) \le 0,\ j \in J\} \ .$$

We assume throughout this section, $f, g_j \in C^2(\mathbb{R}^n \times T, \mathbb{R})$.

We emphazise again, that we have to distinquish between convex and nonconvex programs. If for any $t \in T$ the functions $f(x,t),\ g_j(x,t),\ j \in J$, are convex in x then for any t the program $P(t)$ is convex (cf., Definition 2.5) and any (local) minimizer $x(t)$ is a global one. In this case we can define the (global) value function of $P(t)$ by

$$v(t) = \inf_{x\in F(t)} f(x,t) \ ,$$

($v(t) = \infty$, if $F(t) = \emptyset$) and the set $S(t)$ of global minimizers:

$$S(t) = \{x \in F(t) \mid f(x,t) = v(t)\} \ .$$

In the general nonlinear case, we consider local minimizer $x(t)$ of $P(t)$ and the corresponding local minimum value function,

$$v(t) = f(x(t),t) \ .$$

We extend the notation of Section 2.4 to the parametric context. For $\bar{t} \in T$ and feasible $\overline{x} \in F(\bar{t})$ we denote by $J_0(\overline{x}, \bar{t})$ the active index set,

$$J_0(\overline{x}, \bar{t}) = \{j \in J \mid g_j(\overline{x}, \bar{t}) = 0\} ,$$

and by $L(x, t, \mu)$ the Lagrangean function (near $(\overline{x}, \bar{t})$),

$$L(x, t, \mu) = f(x, t) + \sum_{j \in J_0(\overline{x}, \bar{t})} \mu_j g_j(x, t) .$$

Strict Complementarity (SC), second order conditions (SOC) and the cone of critial directions $C_{\overline{x}, \bar{t}}$ are defined accordingly (cf., Theorem 2.4). To find, near $(\overline{x}, \bar{t})$, local minimizers x of $P(t)$ we are looking for solutions $(x, t, \mu) = (x(t), t, \mu(t))$ of the KKT equations with $\mu_j \geq 0, j \in J_0(\overline{x}, \bar{t})$,

$$\text{(4.13)} \quad H(x, t, \mu) := \begin{array}{r} \nabla_x f(x, t) + \sum\limits_{j \in J_0(\overline{x}, \bar{t})} \mu_j \nabla_x g_j(x, t) = 0, \\ g_j(x, t) = 0, \quad j \in J_0(\overline{x}, \bar{t}) . \end{array}$$

From the sufficient optimality conditions in Theorem 2.4 we obtain the following basic stability result by applying IFT to the KKT-system (4.13).

THEOREM 4.4. *[Local stability result based on IFT]*
Let $\overline{x} \in F(\bar{t})$. Suppose that with multipliers $\overline{\mu}_j$ the KKT condition $\nabla_x L(\overline{x}, \bar{t}, \overline{\mu}) = 0$ is satisfied such that

(1) *LICQ holds at $(\overline{x}, \bar{t})$.*
(2) $\overline{\mu}_j > 0, \ \forall j \in J_0(\overline{x}, \bar{t})$ *(SC)*

and either

(3a) *(order one condition)* $|J_0(\overline{x}, \bar{t})| = n$

or

(3b) *(order two condition)*

$$d^T \nabla_x^2 L(\overline{x}, \bar{t}, \overline{\mu}) d > 0 \ \ \forall d \in T_{\overline{x}, \bar{t}} \setminus \{0\}$$

where $T_{\overline{x}, \bar{t}}$ is the tangent space $T_{\overline{x}, \bar{t}} = \{d \mid \nabla_x g_j(\overline{x}, \bar{t}) d = 0, \ j \in J_0(\overline{x}, \bar{t})\}$.

(According to Theorem 2.4, $\overline{x}$ is a local minimizer of $P(\bar{t})$ of order $s = 1$ in case (3a) and of order $s = 2$ in case (3b).)
Then there exist a neighborhood $B_\varepsilon(\bar{t})$, $\varepsilon > 0$, of $\bar{t}$ and C^1-functions $x : B_\varepsilon(\bar{t}) \to \mathbb{R}^n$, $\mu : B_\varepsilon(\bar{t}) \to \mathbb{R}^{|J_0(\overline{x}, \bar{t})|}$ such that $x(\bar{t}) = \overline{x}$, $\mu(\bar{t}) = \overline{\mu}$ and for any $t \in B_\varepsilon(\bar{t})$ the point $x(t)$ is a strict local minimizer of $P(t)$ (of order 1 in case (3a) and of order 2 in case (3b)) with corresponding multiplier vector $\mu(t)$. Moreover, for $t \in B_\varepsilon(\bar{t})$ the derivative of the value function $v(t) = f(x(t), t)$ is

$$\nabla v(t) = \nabla_t L(x(t), t, \mu(t)) .$$

The derivative of the solution function $x(t)$ is given in case (3a) by

$$\nabla x(t) = -\begin{pmatrix} \nabla_x g_j(x(t),t),\ j \in J_0(\overline{x},\overline{t}) \\ \vdots \end{pmatrix}^{-1} \begin{pmatrix} \nabla_t g_j(x(t),t),\ j \in J_0(\overline{x},\overline{t}) \\ \vdots \end{pmatrix}$$

and in case (3b) by (cf., (4.13))

$$\begin{pmatrix} \nabla x(t) \\ \nabla \mu(t) \end{pmatrix} = -[\nabla_{(x,\mu)} H(x(t),t,\mu(t))]^{-1} \nabla_t H(x(t),t,\mu(t)) \ .$$

Proof. In case (3a), by LICQ and $|J_0(\overline{x},\overline{t})| = n$ the system (4.13) splits into the system of n equations in n variables,

$$g_j(x,t) = 0 \ , \qquad j \in J_0(\overline{x},\overline{t}) \ ,$$

for $x = x(t)$ with $\nabla x(t)$ given as solution of (derivatives wrt. t)

$$\nabla_x g_j(x(t),t)\nabla x(t) + \nabla_t g_j(x(t),t) = 0 \ , \quad j \in J_0(\overline{x},\overline{t}) \ , \tag{4.14}$$

and the equations

$$\nabla_x f(x(t),t) + \sum_{j \in J_0(\overline{x},\overline{t})} \mu_j \nabla_x g_j(x(t),t) = 0 \ , \tag{4.15}$$

for the components $\mu_j = \mu_j(t), j \in J_0(\overline{x},\overline{t})$ ($\mu_j(t) = 0, j \in J \setminus J_0(\overline{x},\overline{t})$). For the value function $v(t) = f(x(t),t)$, by differentiation wrt. t, we find using (4.14) and (4.15)

$$\begin{aligned} \nabla v(t) &= \nabla_x f(x(t),t)\nabla x(t) + \nabla_t f(x(t),t) \\ &= -\sum_{j \in J_0(\overline{x},\overline{t})} \mu_j(t) \nabla_x g_j(x(t),t)\nabla x(t) + \nabla_t f(x(t),t) \\ &= \sum_{j \in J_0(\overline{x},\overline{t})} \mu_j(t) \nabla_t g_j(x(t),t) + \nabla_t f(x(t),t) \ = \ \nabla_t L(x(t),t,\mu(t)) \ . \end{aligned}$$

In case (3b) we have to apply the IFT to the coupled system (4.13) and to make use of Ex. 4.5. to show that the matrix $\nabla_{(x,\mu)} H(\overline{x},\overline{t},\overline{\mu})$ is nonsingular, where

$$\nabla_{(x,\mu)} H(\overline{x},\overline{t},\overline{\mu}) = \begin{pmatrix} \nabla_x^2 L(\overline{x},\overline{t},\overline{\mu}) & B \\ B^T & 0 \end{pmatrix} \text{ with } B = \big(\nabla_x g_j(\overline{x},\overline{t})^T, j \in J_0(\overline{x},\overline{t})\big).$$

$\square$

REMARK 4.1. Note that the statement of Theorem 4.3 for linear programs represents a special case of Theorem 4.4, case (3a).

EX. 4.5. Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and $B \in \mathbb{R}^{n \times m}$ ($n \geq m$). Suppose the matrix B has full rank m and the following holds:

$$d^T A d \neq 0 \quad \forall d \in \mathbb{R}^n \setminus \{0\} \text{ such that} B^T d = 0 \ .$$

Show that then the following matrix is nonsingular:

$$\begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}$$

Ex. 4.6. *Let the KKT conditions (4.13) be satisfied at* $(\overline{x}, \overline{t})$. *Then (see Ex. 2.8)*

$$SC \quad \Rightarrow \quad C_{\overline{x},\overline{t}} = T_{\overline{x},\overline{t}} \ .$$

We add an exercise on a parametric program as in Theorem 4.4 but with only equality constraints which will be used in Chapter 7.

Ex. 4.7. *Consider*

$$P_=(t): \quad \min_{x \in \mathbb{R}^n} f(x,t) \quad \textit{s.t.} \quad x \in F_=(t) = \{x \mid g_j(x,t) = 0 \ , \ j = 1, \ldots, m\} \ ,$$

where $f, g_j \in C^2(\mathbb{R}^n \times T, \mathbb{R})$. *We assume that LICQ holds at* $(\overline{x}, \overline{t})$, $\overline{x} \in F_=(\overline{t})$, *i.e.,* $\nabla_x g_j(\overline{x}, \overline{t})$, $j = 1, \ldots, m$, *are linearly independent, as well as the KKT condition*

$$\nabla_x L(\overline{x}, \overline{t}, \overline{\lambda}) := \nabla f(\overline{x}, \overline{t}) + \sum_{j=1}^m \overline{\lambda}_j \nabla_x g_j(\overline{x}, \overline{t}) = 0$$

with multipliers $\overline{\lambda}_j \in \mathbb{R}$ *(not necessarily* ≥ 0*). Suppose the second order condition holds:*

$$d^T \nabla_x^2 L(\overline{x}, \overline{t}, \overline{\lambda}) d > 0 \quad \forall d \in T^=_{\overline{x},\overline{t}} \setminus \{0\} \ ,$$

where $T^=_{\overline{x},\overline{t}} = \{d \mid \nabla_x g_j(\overline{x}, \overline{t}) d = 0, j = 1, \ldots m\}$. *Then,* $\overline{x}$ *is a local minimizer of* $P_=(\overline{t})$ *of order 2 and there exist a neighborhood* $B_\varepsilon(\overline{t})$, $\varepsilon > 0$, *differentiable functions* $x(t) : B_\varepsilon(\overline{t}) \to \mathbb{R}^n$, $\lambda(t) : B_\varepsilon(\overline{t}) \to \mathbb{R}^m$, $x(\overline{t}) = \overline{x}, \lambda(\overline{t}) = \overline{\lambda}$ *such that for* $t \in B_\varepsilon(\overline{t})$ *the point* $x(t)$ *is a local minimizer of* $P_=(t)$ *and with corresponding multiplier* $\lambda(t)$ *the KKT relation holds with second order condition.*

Proof. For the proof that $\overline{x}$ is local minimizer of $P_=(\overline{t})$ of order 2 we refer to [9, Th.10.6]. To prove the further statements we have to apply the IFT to the system of KKT equations:

$$\begin{aligned} \nabla f(x,t) + \sum_{j=1}^m \lambda_j \nabla_x g_j(x,t) &= 0 \\ g_j(x,t) &= 0, \quad j = 1, \ldots, m \ . \end{aligned}$$

□

Remark. Many further (often difficult) results are dealing with generalizations of these stability results in Theorem 4.4, Ex. 4.7 under weaker assumptions, *i.e.*, LICQ and/or SC does not hold; see *e.g.*, [2], [6] and Chapters 6,7.

The next example presents a monotonicity result needed later on.

Ex. 4.8. *With $g(x,t) = (g_j(x,t), j \in J)$, $(x,t) \in \mathbb{R}^n \times T$, consider the feasible set $F(t)$ of $P(t)$,*

$$F(t) = \{x \mid g(x,t) \le 0\} .$$

Let be given $t_1, t_2 \in T$ and corresponding feasible points $x_1 \in F(t_1),\ x_2 \in F(t_2)$ satisfying with multiplier vectors $\mu_1, \mu_2 \ge 0$ the complementarity conditions

$$\mu_1^T g(x_1, t_1) = 0 \ , \quad \mu_2^T g(x_2, t_2) = 0 \ .$$

Then the monotonicity relation

$$(\mu_2 - \mu_1)^T (g(x_2, t_2) - g(x_1, t_1)) \ge 0$$

holds. In particular this relation is true if $g = g(x)$ does not depend on t.

Proof. In view of the complementarity conditions and $\mu_1^T g(x_2, t_2) \le 0$, $\mu_2^T g(x_1, t_1) \le 0$ we find

$$\begin{aligned}(\mu_2 - \mu_1)^T (g(x_2, t_2) - g(x_1, t_1)) &= \mu_2^T g(x_2, t_2) + \mu_1^T g(x_1, t_1) \\ &\quad -\mu_2^T g(x_1, t_1) - \mu_1^T g(x_2, t_2) \\ &\ge 0 \ .\end{aligned}$$

□

A similar monotonicity property holds for the following right-hand side perturbed convex program,

$$P(t): \quad \min f(x) \quad \text{s.t.} \quad \begin{array}{r} Ax = t \\ x \ge 0 \end{array} \tag{4.16}$$

where f is a C^1 convex function on $\mathbb{R}^n_+$ and $A \in \mathbb{R}^{m \times n}$, $t \in \mathbb{R}^m$. A minimizer $\overline{x} = x(\overline{t})$ of $P(\overline{t})$ satisfies the KKT conditions with multipliers $\overline{\lambda} \in \mathbb{R}^m$, $0 \le \overline{\mu} \in \mathbb{R}^n$ (cf., Remark 2.4):

$$\begin{array}{r} \nabla f(\overline{x})^T + A^T \overline{\lambda} - \overline{\mu} = 0 \\ \overline{\mu}^T \overline{x} = 0 \\ A\overline{x} = \overline{t} \\ \overline{x} \ge 0 \end{array} \tag{4.17}$$

Lemma 4.1. *Let $f \in C^1$ be convex on $\mathbb{R}^n_+$. Let further $\widetilde{x}, \overline{x}$ be (global) minimizers of $P(\widetilde{t}), P(\overline{t})$ with corresponding multipliers $\tilde{\lambda}, \tilde{\mu}$ and $\overline{\lambda}, \overline{\mu}$ in (4.17). Then the monotonicity relation holds:*

$$(\tilde{\lambda} - \overline{\lambda})^T (\widetilde{t} - \overline{t}) \le -\tilde{\mu}^T \overline{x} - \overline{\mu}^T \widetilde{x} \le 0 \ .$$

Proof. By convexity of f (see Ex. 2.2(a)) we have

$$\big(\nabla f(\widetilde{x}) - \nabla f(\overline{x})\big)(\widetilde{x} - \overline{x}) \ge 0 \ .$$

Using the KKT relations for $\widetilde{x}, \overline{x}$ leads to,

$$(\tilde{\mu} - A^T\tilde{\lambda} - \overline{\mu} + A^T\overline{\lambda})^T(\widetilde{x} - \overline{x}) \geq 0 \ ,$$

or in view of $\tilde{\mu}^T\widetilde{x} = \overline{\mu}^T\overline{x} = 0$,

$$\begin{aligned} -(\tilde{\lambda} - \overline{\lambda})^T A(\widetilde{x} - \overline{x}) &= -(\tilde{\lambda} - \overline{\lambda})^T(\widetilde{t} - \overline{t}) \\ &\geq -(\tilde{\mu} - \overline{\mu})^T(\widetilde{x} - \overline{x}) = \tilde{\mu}^T\overline{x} + \overline{\mu}^T\widetilde{x} \geq 0 \ . \end{aligned}$$

Multiplying by -1 proves the statement. $\square$