Lectures on Parametric Optimization: An Introduction

Georg Still, University of Twente

CONTENTS

1.	Parametric equations	3
2.	Parametric unconstrained	
	optimization	7
2.1.	Non-parametric minimization	7
2.2.	Parametric minimization	8
3.	Parametric constraint	
	optimization	11
3.1.	Non-parametric programs	11
3.2.	Parametric programs	15
4.	Linear parametric programs	19
	1	

4.1. Non-parametric linear	
programs	19
4.2. Parametric linear programs	23
5. Applications	30
5.1. Interior point methods	30
5.2. A parametric location	
problem	32
6. Pathfollowing in practice	35
7. General parametric	
programming	37
References	

1. PARAMETRIC EQUATIONS

Consider the non-parametric equation: Solve

$$F(x) := x^2 - 2x - 1 = 0$$

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

(1)
$$F(x, t) := x^2 - 2t^2x - t^4 = 0$$

The solutions are

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

Ex.1 Sketch the solution curve in the (x, t)-space.

At $\overline{t} = 0$ with solution $\overline{x} = 0$ we find for the gradient of *F*

$$\nabla F(\overline{x},\overline{t}) = \begin{pmatrix} 2x - 2t^2 \\ -4tx - 4t^3 \end{pmatrix}_{\overline{x},\overline{t}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \,.$$

<u>Rule.</u> The solution set of F(x, t) = 0 where $F : \mathbb{R}^2 \to \mathbb{R}$ is "normally" (locally) given by a *one-dimensional* solution curve (x(t), t). However at points $(\overline{x}, \overline{t})$ where $\nabla F(\overline{x}, \overline{t}) = 0$ holds the solution set has a **singular**-ity(such as a bifurcation or a nonsmoothness).

4

Two versions of the Implicit Function Theorem

We consider systems of n + p equations in *n* variables:

F(x, t) = 0 where $F : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$. The Implicit Function Theorem (IFT) makes a statement on the structure of the solution set in the "normal" situation. **Theorem 1.** (*IFT for one equation*) Let $F : \mathbb{R}^k \to \mathbb{R}$ be a C^1 -function. Suppose for $\overline{y} \in \mathbb{R}^k$ we have $F(\overline{y}) = 0$ and $\nabla F(\overline{y}) \neq 0$. Then near \overline{y} the solution set $S(F) := \{y \in \mathbb{R}^k \mid F(y) = 0\}$ is a C^1 manifold of dimension k - 1. Moreover at \overline{y}

$$\nabla F(\overline{y}) \perp S(F)$$

and the gradient $\nabla F(\overline{y})$ points into the region where $F(y) > F(\overline{y})$.

Ex.2 Give a geometrical sketch of the statement.

Theorem 2. (General version of the 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 a neighborhood $U_t(\overline{t})$ of \overline{t} the solution set $S(F) := \{(x, t) | F(x, t) = 0\}$ is described by a C^1 -function $x : U_t(\overline{t}) \to \mathbb{R}$ such that $x(\overline{t}) = \overline{x}$ and

6

F(x(t), t) = 0 for $t \in U_t(\overline{t})$.

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

 $\nabla x(t) = -[\nabla_x F(x(t), t)]^{-1} \nabla_t F(x(t), t) .$

Proof. See *e.g.*, [7]. Note that if x(t) is a C^1 -function satisfying F(x(t), t) = 0 then by differentiation wrt. *t* we find

 $\nabla_{x}F(x(t), t)\nabla x + \nabla_{t}F(x(t), t) = 0.$

2. PARAMETRIC UNCONSTRAINED OPTIMIZATION

2.1. Non-parametric minimization. We assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2 -function. The point $\overline{x} \in \mathbb{R}^n$ is a *local minimizer* of f if there is some $\varepsilon > 0$ such that

$$\begin{split} f(\overline{x}) &\leq f(x) & \forall x, \ \|x - \overline{x}\| < \varepsilon \ . \\ \text{It is called a$$
strict $local minimizer if:} \\ f(\overline{x}) &< f(x) \ \forall x \neq \overline{x}, \ \|x - \overline{x}\| < \varepsilon . \end{split}$

Theorem 3. (Necessary and sufficient optimality conditions) (see e.g., [4]) (a) If \overline{x} is a local minimizer of f then $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x}) \ge 0$ (pos. semidef.) (b) If \overline{x} satisfies $\nabla f(\overline{x}) = 0$ and $\nabla^2 f(\overline{x}) > 0$ (pos. definite) then \overline{x} is a strict local minimizer of f. 2.2. **Parametric minimization.** Let f(x, t)be a C^2 function, $f : \mathbb{R}^n \times T \to \mathbb{R}$, where $T \subset \mathbb{R}^p$ is open. We consider the *parametric problem*: for $t \in T$ find a (local) solution x = x(t) of

(2)
$$P(t)$$
: $\min_{x \in \mathbb{R}^n} f(x, t)$

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

(3)
$$F(x, t) := \nabla_x f(x, t) = 0$$
.

The next examples show the possible bad behavior.

Ex.3 For

8

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

the minimizer is given by x(t) = |t| with minimal value $v(t) := f(x(t), t) = -\frac{2}{3}|t|^3$. Ex.4 For

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

the critical points are given by the curves $x_{1,2}(t) = t^2(1 \pm \sqrt{2})$ and the minimizer by $x_1(t)$.

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

- The value function v(t) = f(x(t), t)may behave "smoother" than the minimizer function x(t).
- A singular behavior appears 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.5 Check the singular behavior for the examples Ex.3 and Ex.4.

The next theorem describes the situation near a non-singular solution $(\overline{x}, \overline{t})$ of (3).

Theorem 4. (local stability result) Let \overline{x} be a solution of $P(\overline{t})$, $\overline{t} \in T$, such that $\nabla_x f(\overline{x}, \overline{t}) = 0$ and $\nabla_x^2 f(x, t) > 0$ (pos. def.). Then in a neighborhood $U_t(\overline{t})$ there is a C^1 -function $x : U_t(\overline{t}) \to \mathbb{R}^n$ such that $x(\overline{t}) =$ \overline{x} and for any $t \in U_t(\overline{t})$, x(t) is a strict local minimizer of P(t). Moreover for $t \in U_t(\overline{t})$, $\nabla x(t) = -[\nabla_x^2 f(x(t), t)]^{-1} \nabla_{xt}^2 f(x(t), t)$,

and the value function v(t) := f(x(t), t)is a C²-function with

$$\nabla v(t) = \nabla_t f(x(t), t)$$

and

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

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

Rem. Note that x(t) is C^1 but v(t) is C^2 .

3. PARAMETRIC CONSTRAINT OPTIMIZATION

3.1. Non-parametric programs. Consider *nonlinear programs*:

(4)

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

with index set $J = \{1, ..., m\}$. The set \mathcal{F} is called *feasible set*.

Def. A point $\overline{x} \in \mathcal{F}$ is called local minimizer of order s = 1 or s = 2 if there are constants $c, \varepsilon > 0$ such that

$$f(x) - f(\overline{x}) \ge c \|x - \overline{x}\|^s \quad \forall x \in \mathcal{F}, \|x - \overline{x}\| < \varepsilon$$

It is a *global* minimizer if $f(x) \ge f(\overline{x})$ holds $\forall x \in \mathcal{F}$.

For $\overline{x} \in \mathcal{F}$ we introduce the *active index set*

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

and the Lagrangian function

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

The coefficients μ_j are called *Lagrangian* multipliers. We say that the *Linear Inde*pendence constraint qualification (LICQ) is satisfied at $\overline{x} \in \mathcal{F}$ if

 $\nabla g_j(\overline{x}), j \in J_0(\overline{x})$, are linearly independent. The next theorem gives the famous *Karush-Kuhn-Tucker* (KKT) sufficient optimality conditions. **Theorem 5.** (Sufficient optimality conditions) Let $\overline{x} \in \mathcal{F}$ satisfy LICQ. (a) (Order one) Let with multipliers $\overline{\mu}_j$ the KKT condition (5) $\nabla_x L(\overline{x}, \overline{\mu}) = \nabla f(\overline{x}) + \sum_{j \in J_0(\overline{x})} \overline{\mu}_j \nabla g_j(\overline{x}) = 0$, $\overline{\mu}_j \ge 0, \ j \in J_0(\overline{x}), \ be \ satisfied \ such \ that$ $\overline{\mu}_j > 0, \ \forall j \in J_0(\overline{x}) \ (Strict \ complement. \ (SC))$ and $|J_0(\overline{x})| = n$. Then \overline{x} is a local minimizer of (P) of order s = 1.

13

14

(**b**) (Order two) Let with multipliers $\overline{\mu}_j$ the KKT condition (5) be satisfied such that (SC) holds and the second order condition (SOC)

SOC: $d^T \nabla_x^2 L(\overline{x}, \overline{\mu}) d > 0 \ \forall d \in T_{\overline{x}} \setminus \{0\}$ where $T_{\overline{x}}$ is the tangentspace $T_{\overline{x}} = \{d \mid \nabla g_j(\overline{x}) d = 0, \ j \in J_0(\overline{x})\}$. Then \overline{x} is a local minim. of (P) of order s = 2.

Ex.6 Show that the point $\overline{x} = 0$ is the minimizer of order s = 1 of the problem $\min x_2$ s.t. $e^{-x_1} - x_2 - 1 \le 0$, $x_1 - x_2 \le 0$

Rem. The KKT conditions can also be given in the equivalent (global) form:

(6)
$$\nabla f(x) + \sum_{j \in J} \mu_j \nabla g_j(x) = 0$$

 $\mu_j \cdot g_j(x) = 0, \ j \in J$
 $\mu_j, -g_j(x) \ge 0, \ j \in J.$

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

(7)
$$P(t): \min_{x \in \mathbb{R}^n} f(x, t)$$
 s.t. $x \in \mathcal{F}(t)$,

where

$$\mathcal{F}(t) := \{ x \mid g_j(x, t) \le 0 \ , \ j \in J \} \ .$$

For $\overline{t} \in T$ and feasible $\overline{x} \in \mathcal{F}(\overline{t})$ we denote by $J_0(\overline{x}, \overline{t})$ the active index set and by $L(x, t, \mu)$ the Lagrangian function

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

To find (near $(\overline{x}, \overline{t})$) local minimizers x of P(t) we are looking for solutions (x, t, μ) of the KKT-equations (with $\mu_j \ge 0$) (8)

$$F(x, t, \mu) := \frac{\nabla_x L(x, t, \mu) = 0}{g_j(x, t) = 0}, j \in J_0(\overline{x}, \overline{t})$$

16

From the sufficient optimality conditions in Theorem 5 we obtain

Theorem 6. (Local stability result) Let $\overline{x} \in \mathcal{F}(\overline{t})$. Suppose that with multipliers $\overline{\mu}_j$ the KKT condition $\nabla_x L(\overline{x}, \overline{t}, \overline{\mu}) = 0$ is satisfied such that

(1) LICQ holds at \overline{x} wrt. $\mathcal{F}(\overline{t})$. (2) $\overline{x} \ge 0$ $\forall i \in L(\overline{x}, \overline{t})$ (54)

(2) $\overline{\mu}_j > 0, \ \forall j \in J_0(\overline{x}, \overline{t})$ (SC) and either

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

or

(3b) (order two)

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

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

(According to Theorem 5, \overline{x} is a local minimizer of $P(\overline{t})$ of order s = 1 in case (3a) and of order s = 2 in case (3b)). Then there exist a neighborhood $U_t(\bar{t})$ of \bar{t} and C^1 -functions $x : U_t(\bar{t}) \to \mathbb{R}^n$, $\mu :$ $U_t(\bar{t}) \to \mathbb{R}^{|J_0(\bar{x},\bar{t})|}$ such that $x(\bar{t}) = \bar{x}$, $\mu(\bar{t}) =$ $\bar{\mu}$ and for any $t \in U_t(\bar{t})$ the point x(t) is a strict local minimizer of P(t) with corresponding multiplier $\mu(t)$. Moreover for $t \in U_t(\bar{t})$ the derivatives of x(t), $\mu(t)$ and the value function v(t) = f(x(t), t) are given by

17

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

and

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

18

Ex.7 Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix and $B \in \mathbb{R}^{n \times m}$ $(n \ge m)$. Suppose the matrix *B* has full rank *m* and the following holds:

 $d^T A d \neq 0 \quad \forall d \in \mathbb{R}^n$ such that $B^T d = 0$. Show that then the following matrix is regular:

 $\left(\begin{array}{cc} A & B \\ B^T & 0 \end{array}\right)$

Rem. Many further (often difficult) results are dealing with the generalization of this stability result under weaker assumptions (*i.e.*, LICQ or SC does not hold; see *e.g.*, [2], [3]).

4. LINEAR PARAMETRIC PROGRAMS

4.1. Non-parametric linear programs. Let be given a matrix $A \in \mathbb{R}^{m \times n}$ with *m* rows a_j^T , $j \in J := \{1, \ldots, m\}$ and vectors $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$. We consider *primal* problems of the form

(9)
$$P: \max c^T x \text{ s.t. } x \in \mathcal{F}_P$$

$$\mathcal{F}_P = \{x \mid a_j^T x \le b_j, \ j \in J\}.$$

We often write the feasible set in compact form $\mathcal{F} = \{x \mid Ax \leq b\}$. The problem

(10)
$$D: \min b^T y \text{ s.t. } y \in \mathcal{F}_D$$

$$\mathcal{F}_D = \{ y \mid \sum_{j=1}^m y_j a_j = c, y \ge 0 \}$$

is called the dual problem. A vector $x \in \mathcal{F}_P$ is called feasible for *P* and a vector *y* satisfying the feasibility conditions $A^T y = c, y \ge 0$ is called feasible for *D*. Let v_P ,

 v_D resp. denote the maximum value of P resp. minimum value of D.

Lemma 1. (weak duality) Let x be feasible for P and y be feasible for D. Then $c^T x \le b^T y$ and thus $v_P \le v_D$. If $c^T x = b^T y$ holds then x is a maximizer of P and y a minimizer of D.

Proof. As Ex.

For $\overline{x} \in \mathcal{F}_P$ as usual we define the active index set $J_0(\overline{x}) = \{j \mid a_j^T x = b_j\}$. For a subset $J_0 \subset J$ we denote by A_{J_0} the submatrix of A with rows a_j^T , $j \in J_0$ and for $y \in \mathbb{R}^m$ by y_{J_0} the vector $(y_j, j \in J_0)$.

Def. A feasible point $\overline{x} \in \mathcal{F}_P$ is called a vertex of the polyhedron \mathcal{F}_P if the vectors

 $a_j, j \in J_0(\overline{x})$ form a basis of \mathbb{R}^n or equivalently if $A_{J_0(\overline{x})}$ has rank *n*. (This implies $|J_0(\overline{x})| \ge n$. The vertex \overline{x} is called *nondegenerate* if LICQ holds, *i.e.*, a_j , $j \in J_0(\overline{x})$ are linear independent (implying $|J_0(\overline{x})| = n$ or equivalently $A_{J_0(\overline{x})}$ is non-singular).

Theorem 7.

(a) (Existence and strong duality) If both P and D are feasible then there exist solutions \overline{x} of P and \overline{y} of D. Moreover for (any of) these solutions we have

 $c^T \overline{x} = b^T \overline{y}$ and thus $v_P = v_D$.

(b) (Optimality conditions) A point $\overline{x} \in \mathcal{F}_P$ is a solution of P if and only if there is a corresponding $\overline{y} \in \mathcal{F}_D$ such that the complementarity conditions

$$\overline{y}^T(b - A\overline{x}) = 0$$
 or $\overline{y}_j(b_j - a_j^T\overline{x}) = 0, \forall j \in J$

hold or equivalently if there exists \overline{y}_j , $j \in J_0(\overline{x})$, such that KKT conditions are satisfied:

22

 $\sum_{j \in J_0(\overline{x})} \overline{y}_j a_j = c, \quad \overline{y}_j \ge 0, \ j \in J_0(\overline{x}) \ .$

It appears that normally the solution of P arises at a vertex of \mathcal{F}_P .

Lemma 2. If the polyhedron \mathcal{F}_P has a vertex (at least one) and $v_P < \infty$ then the max value v_P of P is also attained at some vertex of \mathcal{F}_P .

Rem. The Simplex algorithm for solving *P* proceeds from vertex to vertex of F_P until the optimality conditions in Theorem 7(b) are met.

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

(11) P(t): max $c^T(t)x$ s.t. $x \in \mathcal{F}_P(t)$ $\mathcal{F}_P(t) = \{x \mid A(t)x \le b(t)\}$

The corresponding dual reads (12) D(t): min $b^T(t)y$ s.t. $y \in \mathcal{F}_D$ $\mathcal{F}_D = \{y \mid A^T(t)y = c(t), y \ge 0\}$. For $\overline{t} \in T$ and $\overline{x} \in \mathcal{F}_P(\overline{t})$ the active index

set is $J_0(\overline{x}, \overline{t})$.

Linear Production Model, Shadow Prices 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_j requires a_{ij} units of resource R_i , for $i = 1, \ldots, m$.

Suppose we can sell our production for the price of c_j per 1 unit of P_j and that b_i units of each resource R_i are available for the total production. How many units x_j of each product P_j should we produce in order to maximize the total receipt from the sales?

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

(13)

 $P: \max c^T x \quad \text{s.t.} \quad Ax \le b, \ x \ge 0.$

24

Here $A = (a_{ij})$ is the matrix with the elements a_{ij} . Let \overline{x} be a solution with corresponding solution \overline{y} of the dual problem (14)

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

and maximum profit $\overline{z} = c^T \overline{x} = b^T \overline{y}$.

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_i ?

Let us increase (for fixed *i*) the capacity of R_i from b_i to $b'_i = b_i + t$, \bar{y} is still feasible. So weak duality gives the following upper bound on the expected profit:

$$\overline{z}' \leq b_1 \overline{y}_1 + \dots + b'_i \overline{y}_i + \dots + b_m \overline{y}_m$$
$$= t \cdot \overline{y}_i + \sum_{s=1}^m b_s \overline{y}_s = t \cdot \overline{y}_i + \overline{z} .$$

Accordingly, we would not want to pay more than $t \cdot \bar{y}_i$ for t more units of R_i . In this sense, the coefficients \bar{y}_i of the dual optimal solution \bar{y} can be interpreted as the *shadow prices* of the resources R_i .

Note that for the value function v(t) in dependence from the parameter *t* we find

$$\frac{v(t) - v(0)}{t} = \frac{\overline{z}' - \overline{z}}{t} \le \overline{y}_i .$$

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} , 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.

26

General case. We come back to the general parametric LP in (11). Suppose for $\overline{t} \in T$ the point $\overline{x} \in \mathcal{F}_P(\overline{x})$ 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:

(15)
$$P: \quad A_{J_0(\overline{x},\overline{t})}x = b_{J_0(\overline{x},\overline{t})}$$

and

(16)

$$D: \quad A_{J_0(\overline{x},\overline{t})}^T y_{J_0(\overline{x},\overline{t})} = c_{J_0(\overline{x},\overline{t})}, \quad y_{J_0(\overline{x},\overline{t})} \ge 0$$

If \overline{x} is a non-degenerate vertex, *i.e.*, $A_{J_0(\overline{x},\overline{t})}$ is nonsingular, this is possible by applying the IFT to these systems.

Theorem 8. (Local stability result) Let $\overline{x} \in \mathcal{F}_P(\overline{t})$ be a vertex solution 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 5 \overline{x} is a local minimizer of $P(\overline{t})$ of order s = 1.) Then there exist a neighborhood $U_t(\overline{t})$ and C^1 -functions $x : U_t(\overline{t}) \to \mathbb{R}^n$, $y_{J_0(\overline{x},\overline{t})} : U_t(\overline{t}) \to \mathbb{R}^{|J_0(\overline{x},\overline{t})|}$ 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 $t \in U_t(\overline{t})$ the point x(t) is a vertex solution of P(t) (of order s=1) with corresponding multiplier $y_{J_0(\overline{x},\overline{t})}(t)$. Moreover for $t \in U_t(\overline{t})$ the derivatives of x(t)and the value function $v(t) = c^T(t)x(t)$

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

and
$$\nabla v(t) = [\nabla c_{J_0(\overline{x},\overline{t})}(t)]^T x$$

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

Rem. The production model case above is a special case where *A*, *c* do not depend on $t \in \mathbb{R}$ (so p = 1) and (for fixed $i \in J_0(\overline{x}, \overline{t})$) $\frac{1}{dt}b_{J_0(\overline{x}, \overline{t})}(t) = e_i$ (e_i is the *i*th unit vector in $\mathbb{R}^{|J_0(\overline{x}, \overline{t})|}$) the so that

$$\frac{1}{dt}v(\overline{t}) = [y_{J_0(\overline{x},\overline{t})}(\overline{t})]^T e_i = \overline{y}_i.$$

5. APPLICATIONS

5.1. **Interior point methods.** The basic idea of the interior point method for solving a non-parametric program

 $P: \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_j(x) \le 0, \ j \in J$ is simply as follows. Consider the perturbed KKT system (17) $\nabla f(x) + \sum_{j \in J} \mu_j \nabla g_j(x) = 0$

 $F(x, \tau, \mu) := \begin{array}{c} j \in J \\ -\mu_j \cdot g_j(x) = \tau, \ j \in J, \\ \mu_j, \ -g_j(x) \ge 0, \ j \in J. \end{array}$

where $\tau > 0$ is a perturbation parameter. The idea is to find solutions $x(\tau)$ and $\mu_j(\tau)$ of this system (satisfying $-g_j(x(\tau)), \mu_j(\tau) >$ 0) and to let $\tau \downarrow 0$. We expect that $x(\tau)$ converges to a solution \overline{x} of *P*. Under the sufficient optimality conditions this procedure is well-defined (at least for small τ).

Theorem 9. Let \overline{x} be a local minimizer of P such that the sufficient optimality conditions of Theorem 5 are fulfilled with multiplier $\overline{\mu}$ so that $\overline{\mu}_j > 0$ holds for all $j \in J_0(\overline{x})$. Then there exists C^1 -functions $x : (-\alpha, \alpha) \to \mathbb{R}^n$, $\mu_j : (-\alpha, \alpha) \to \mathbb{R}$, $j \in J_0(\overline{x})$ ($\alpha > 0$) such that $x(0) = \overline{x}$, $\mu_j(0) = \overline{\mu}_j$, and $x(\tau)$, $\mu_j(\tau)$ are locally unique solutions of (17)

Proof. Follows by applying the IFT to the equation (17). \Box

5.2. A parametric location problem. We consider a concrete location problem in Germany (see [6] for details). Suppose some good is produced at 5 existing plants at location $s^{j} = (s_{1}^{j}, s_{2}^{j}), j = 1, ..., 5 (s_{1}^{j})$ longitude, s_2^J latitude in Germany) and a sixth new plant has to be build at location $t = (t_1, t_2)$ (to be determined) to satisfy the demands of V_i units of goods in 99 towns *i* at location $\ell^i = (\ell_1^i, \ell_2^i), i =$ $1, \ldots, 99$. Suppose (for simplicity) that the transportation cost c_{ij} from plant j to town *i* (per unit of the good) is (proportionally to) the Euclidian distance

$$c_{ij} = \sqrt{(s_1^j - \ell_1^i)^2 + (s_2^j - \ell_2^i)^2}, \quad j = 1, ..., 5,$$

$$c_{i6}(t) = \sqrt{(t_1 - \ell_1^i)^2 + (t_2 - \ell_2^i)^2}, \quad \forall i.$$

Suppose further that the total demand $V = \sum_i V_i$ will be produced in the 6 plants with

a production of p_j units of the good in plant j where

$$p_1 = V \frac{10}{100}, \quad p_2 = V \frac{30}{100}, \quad p_3 = V \frac{10}{100},$$
$$p_4 = V \frac{15}{100}, \quad = p_5 = V \frac{15}{100}, \quad p_6 = V \frac{20}{100}.$$
The problem now is to find the location $t = (t_1, t_2)$ of the new plant such that the total transportation costs are minimized. For any fixed location t the optimal transportation strategy is given by the solution of the transportation problem (standard LP),

$$P(t) ::= \min_{y_{ij}} \sum_{j=1}^{5} \sum_{i} c_{ij} y_{ij} + \sum_{i} c_{i6}(t) y_{i6} \quad \text{s.t.}$$
$$\sum_{i} y_{ij} = p_{j}, \quad j = 1, ..., 6,$$
$$\sum_{j} y_{ij} = V_{i}, \quad i = 1, ..., 99.$$

Here y_{ij} is the number of units of the good to be transported from plant *j* to town *i*. The problem is now to find a (global or local) minimizer of v(t). Most local minimization algorithm are based on the computation of the (negative) gradient of the objective function at some actual point $t = t^k$, $d = -\nabla v(t^k)$.

Suppose that y_{ij}^k is the vertex solution of $P(t^k)$. Then by the results of Section 4.2 the gradient can be computed via the formula

$$\nabla v(t^k) = \nabla_t L(y^k, t^k, \lambda^k) =$$

$$\sum_{i=1}^{99} y_{i6}^k \frac{1}{\sqrt{(t_1 - \ell_1^i)^2 + (t_2 - \ell_2^i)^2}} \begin{pmatrix} (t_1 - \ell_1^i) \\ (t_2 - \ell_2^i) \end{pmatrix}$$

6. PATHFOLLOWING IN PRACTICE

We shortly discuss how a solution curve (x(t), t) of a one-parametric equation F(x, t) = 0 $F: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n, t \in \mathbb{R},$ 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:

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

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

 $x^{k+1} = x^k - [\nabla F(x^k)]^{-1} F(x^k), \quad k = 0, 1, \dots$ It is well-known that this iteration converges quadratically to a solution \overline{x} of $F(\overline{x}) = 0$ if

35

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

The simplest way to follow approximately a solution curve x(t) of F(x, t) = 0, *i.e.*, F(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 $\ell = 0, ..., N$, a solution $x_{\ell} = x(t_{\ell})$ of $F(x, t_{\ell}) = 0$ by a Newton iteration, $x_{\ell}^{k+1} = x_{\ell}^{k} - [\nabla F(x_{\ell}^{k}, t_{\ell})]^{-1} F(x_{\ell}^{k}, t_{\ell}), \quad k = 0, 1, ...$ starting with $x_{\ell}^{0} = x_{\ell-1}$ (for $\ell \geq 1$).

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?

7. GENERAL PARAMETRIC PROGRAMMING

In the next sections we analyze the parametric behavior under weaker assumptions where the Implicit Function Theorem is no more applicable. We try to keep the introduction of 'new' concepts to a minimum and to motivate the results by examples.

Consider again the parametric optimization problem

(18) $P(t): \min_{x \in \mathbb{R}^n} f(x, t)$ s.t. $x \in F(t)$ $F(t):= \{x \in \mathbb{R}^n \mid g_j(x, t) \le 0, j \in J\},$ depending on the parameter $t \in T$, where $T \subset \mathbb{R}^p$ is an open parameter set. Again $J := \{1, \ldots, m\}.$ All functions f, g_j are assumed to be (at least) continuous everywhere. **Notation:** Let $v(t) = \min_{x \in \mathcal{F}(t)} f(x, t)$ denote the minimal value of P(t) ($v(t) = \infty$ if $F(t) = \emptyset$) and let S(t) the set of (global) minimizers. The mappings F: $T \Rightarrow \mathbb{R}^n$ and $S: T \Rightarrow \mathbb{R}^n$ are so-called setvalued mappings.

38

Problem of parametric optimization: How do the value function v(t) and the mappings F(t), S(t) change with t. (Continuously, smoothly?)

Definition 1. Let $v : T \to \mathbb{R}_{\infty}$ be given, $\mathbb{R}_{\infty} = \mathbb{R} \cup \{-\infty, \infty\}.$

(a) The function v is called *upper semi*continuous (usc) at $\overline{t} \in T$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

 $v(t) \le v(\overline{t}) + \varepsilon$ for all $||t - \overline{t}|| < \delta$.

(b) The function v is called *lower semi*continuous (lsc) at $\overline{t} \in T$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$v(t) \ge v(\overline{t}) - \varepsilon$$
 for all $||t - \overline{t}|| < \delta$.

We shall see that the lower- and upper semicontinuity of the value function v(t)depend on different assumptions. Obviously to assure the lower semicontinuity of v at \bar{t} the feasible set F(t) should not become essentially larger by a small perturbation t of \bar{t} and to assure the upper semicontinuity of v the set F(t) should not become essentially smaller after a perturbation. To avoid an 'explosion' of F(t)we will need some compactness assumptions for F(t) and to prevent an 'implosion' a Constraint Qualification will be needed. **Definition 2.** Let the set valued mapping $F: T \rightrightarrows \mathbb{R}^n$ be given.

(a) *F* is called *closed* at $\overline{t} \in T$ if for any sequences $t_l, x_l, l \in \mathbb{N}$, with $t_l \to \overline{t}, x_l \in F(t_l)$ the condition $x_l \to \overline{x}$ implies $\overline{x} \in F(\overline{t})$.

(**b**) (no explosion of F(t) after perturbation of $t = \overline{t}$) F is called outer semicontinuous (osc) at $\overline{t} \in T$ if for any sequences $t_l, x_l, l \in \mathbb{N}$ with $t_l \to \overline{t}, x_l \in F(t_l)$ there exists $\overline{x}_l \in F(\overline{t})$ such that $||x_l - \overline{x}_l|| \to 0$ for $l \to \infty$.

(c) (no implosion of F(t) after perturbation of $t = \overline{t}$) F is inner semicontinuous (isc) at $\overline{t} \in T$ if for any $\overline{x} \in F(\overline{t})$ and sequence $t_l \to \overline{t}$ there exists a sequence x_l such that $x_l \in F(t_l)$, for l large enough, and $||x_l - \overline{x}|| \to 0$.

40

The mapping *F* is called continuous at \overline{t} if it is both osc and isc at \overline{t}

Ex.8 Show for $F(t) = \{x \in \mathbb{R}^n \mid g_j(x, t) \le 0, j \in J\}, t \in T (g_j \text{ continuous}) \text{ that the mapping } F: T \Longrightarrow \mathbb{R}^n \text{ is closed on } T.$

Lower semicontinuity of v(t). To assure lower semicontinuity of v at \overline{t} we (minimally) need compactness of $F(\overline{t})$ (even in the case that F(t) behaves continuously). **Ex.9** (Linear problem, $F(t) \equiv F$ constant not bounded and v is not lsc.) For the problem min $x_2 - tx_1$ s.t. $x_1 \ge 0$, $x_2 \ge 0$ we find

$$v(t) = \begin{cases} 0 & \text{for } t \le 0 \\ -\infty & \text{for } t > 0 \end{cases} \text{ and}$$
$$S(t) = \begin{cases} \{(0,0)\} & \text{for } t < 0 \\ \{(x_1,0) \mid x_1 \ge 0\} & \text{for } t = 0 \\ \emptyset & \text{for } t > 0 \end{cases}$$

We even need the following stronger condition

LC. (*local compactness* of *F* at \overline{t}) There exists $\varepsilon > 0$ and a compact set C_0 such that

$$\bigcup_{\|t-\bar{t}\|\leq\varepsilon}F(t)\subset C_0.$$

Without this condition LC the lower semicontinuity of v is not assured in general. **Ex.10** ($F(\bar{t})$, $S(\bar{t})$ compact, LC does not hold and v is not lsc and F(t) is not osc at \bar{t} .) Consider the problem

min
$$x_2 - x_1$$
 s.t. $x_2 \le 2tx_1 - \frac{1}{2}$,

 $x_2 \leq -tx_1, \ x_2 \geq x_1(t)$,

with a function $x_1(t)$, $x_1(0) = -1$, $x_1(t) = -\frac{1}{4}$ for $|t| \ge 2$ (sketch the problem as an Ex.) we find with continuous functions

42

$$v_{l}(t), x(t):$$

$$v(t) = \begin{cases} v_{l}(t) & \text{for } t \leq 0 \\ -\frac{1}{2} - \sqrt{\ln(2)} & \text{for } t = 0 \\ -\frac{1}{4}(1 + \frac{1}{t}) & \text{for } t > 0 \end{cases}$$

$$S(t) = \begin{cases} \{(x(t))\} & \text{for } t \le 0\\ \{\frac{1}{4}(\frac{1}{t}, -1)\} & \text{for } t > 0 \end{cases}.$$

The lower semicontinuity of v depends on the following technical outer semicontinuity and compactness condition saying that for t near \overline{t} at least one point $x_t \in S(t)$ can be approached by elements in a compact subset of $F(\overline{t})$.

AL. There exists a neighborhood $U_t(\overline{t})$ of \overline{t} and a compact set $\overline{C} \subset \mathbb{R}^n$ such that for all $t \in U_t(\overline{t})$ with $F(t) \neq \emptyset$ there is some $x_t \in S(t)$ and some $\overline{x}_t \in F(\overline{t}) \cap \overline{C}$ satisfying $||x_t - \overline{x}_t|| \to 0$ for $t \to \overline{t}$.

and

Lemma 3.

44

(a) Let AL be satisfied at \overline{t} . Then v is lsc at \overline{t} .

(b) Let the local compactness condition *LC* be fulfilled at \overline{t} . Then AL is satisfied (i.e., v is lsc at \overline{t}).

Ex.11 Let the local compactness condition LC be fulfilled at \overline{t} then the mapping *F* is osc at \overline{t} .

The convex case. By Ex.10 in the general (non-convex) case the condition $\emptyset \neq F(\overline{t})$ compact is not sufficient to assure the outer semicontinuity of *F* and the condition $\emptyset \neq S(\overline{t})$ compact does not imply the lower semicontinuity of *v*. Under the following convexity assumptions the situation is changed.

Recall that a function f(x), $f : \mathbb{R}^n \to \mathbb{R}$ is called convex if for any $x_1, x_2 \in \mathbb{R}^n$ and $\tau \in [0, 1]$ it follows

 $f((1-\tau)x_1 + \tau x_2) \le (1-\tau)f(x_1) + \tau f(x_2) .$

AC. For each (fixed) $t \in T$ the functions $g_j(x, t)$ are convex in *x* for all $j \in J$.

AC₀. In addition to AC, for each (fixed) $t \in T$ the function f(x, t) is convex in x, (*i.e.*, the problems P(t) are convex programs).

Lemma 4. Let the convexity condition AC hold and assume $\emptyset \neq F(\overline{t})$ is bounded (compact). Then the local compactness condition LC is satisfied. In particular F is osc (cf. Ex.11) and v is lsc at \overline{t} (cf. Lemma 3).

In the convex case even the boundedness of $S(\overline{t})$ is sufficient to assure that v is lsc at \overline{t} . 46

Theorem 10. Let the convexity assumption AC_0 hold and let $S(\overline{t})$ be nonempty and bounded (compact). Then v is lsc at \overline{t} .

The upper semicontinuity of v(t). The upper semicontinuity of v depends on a different assumption. The local compactness condition LC is not sufficient (and not necessary).

Ex.12 For the problem

min x_1 s.t. $x_1^2 + x_2^2 \le -t$, which obviously satisfies LC we find

$$F(t) = \begin{cases} \{x_1^2 + x_2^2 \le |t|\} \text{ for } t < 0\\ \{(0,0)\} \text{ for } t = 0\\ \emptyset \text{ for } t > 0 \end{cases} \text{ and }$$

$$v(t) = \begin{cases} t & \text{for } t \le 0\\ \infty & \text{for } t > 0 \end{cases}$$

The problem here is that the mapping F is not isc at \overline{t} . To assure the upper semicontinuity of v at \overline{t} we only need an inner semicontinuity condition at one point $\overline{x} \in S(\overline{t})$.

AU. There exist a minimizer $\overline{x} \in S(\overline{t})$ and a neighborhood $U_t(\overline{t})$ of \overline{t} such that for all $t \in U_t(\overline{t})$ there is some $x_t \in F(t)$ satisfying $||x_t - \overline{x}|| \to 0$ for $t \to \overline{t}$.

Lemma 5. Let AU be fulfilled at \overline{t} . Then v is usc at \overline{t} .

A natural condition to force the assumption AU is a so-called *Constraint Qualification* (CQ).

Definition 3. *The* Constraint Qualification *CQ* is said to hold at $(\overline{x}, \overline{t})$ with $\overline{x} \in F(\overline{t})$ if there is a sequence $x_{\nu} \to \overline{x}$ such that $g_j(x_{\nu}, \overline{t}) < 0$ holds for all $j \in J$. **Corollary 1.** Let CQ be satisfied at $(\overline{x}, \overline{t})$ for (at least) one point $\overline{x} \in S(\overline{t})$. Then the condition AU is satisfied, i.e., v is usc at \overline{t} .

48

Ex.13 Let CQ hold at $(\overline{x}, \overline{t}), \overline{x} \in F(\overline{t})$. Then there is a neighborhood $U_t(\overline{t})$ of \overline{t} such that F(t) is nonempty for all $t \in U_t(\overline{t})$. If CQ holds at (x, \overline{t}) for each point $x \in F(\overline{t})$ then *F* is isc at \overline{t} .

The behavior of S(t). Let us shortly study the continuity properties of the mapping S(t).

Obviously the inner semicontinuity of S is stronger than the condition AU. So if S is isc the function v is usc (see Lemma 5). However the inner semicontinuity of S(t) is a very 'strong' condition. Even if the local compactness condition LC and the Constraint Qualification hold it need not to be satisfied. We give an example.

Ex. 14 (LC and CQ holds but *S* is not isc.) min $x_2 - tx_1$ s.t. $|x_1| \le 1$, $|x_2| \le 1$. Then near $\overline{t} = 0$ we obtain

$$S(t) = \begin{cases} \{(-1, -1)\} & \text{for } t < 0\\ \{(x_1, -1) \mid |x_1| \le 1\} & \text{for } t = 0\\ \{(1, -1)\} & \text{for } t > 0 \end{cases}$$
$$v(t) = -1 - |t|.$$

We now discuss the closedness and the outer semicontinuity of *S*.

Ex.15 Let CQ be satisfied at $(\overline{x}, \overline{t})$ for (at least) one point $\overline{x} \in S(\overline{t})$. Then S is closed at \overline{t}

Ex. 16 Let $S(\overline{t})$ be compact and assume $\emptyset \neq S(t)$ for a neighborhood $U_t(\overline{t})$ of \overline{t} and S is osc at \overline{t} . Then AL holds implying that v is lsc at \overline{t} (see Lemma 3).

49

and

Lemma 6. Let LC be satisfied at \overline{t} and let CQ hold at $(\overline{x}, \overline{t})$ for (at least) one $\overline{x} \in S(\overline{t})$. Then there exists a neighborhood $U_t(\overline{t})$ of \overline{t} such that for all $t \in U_t(\overline{t})$ the set S(t) is nonempty and compact. Moreover the mapping S is osc at \overline{t} .

REFERENCES

- [1] Allgower E.L., Georg K, Numerical Continuation Methods., Springer-Verlag, Berlin, (1990).
- [2] Bank G., Guddat J., Klatte D., Kummer B., Tammer D., *Non-linear Parametric Optimization*, Birkhäuser Verlag, Basel, (1983).
- [3] Bonnans F., Shapiro, Alexander Perturbation analysis of optimization problems. Springer Series in Operations Research. Springer-Verlag, New York, (2000).
- [4] Faigle U., Kern W., Still G., Algorithmic Principles of Mathematical Programming, Kluwer, Dordrecht, (2002).
- [5] Guddat J., Guerra F. and Jongen H. Th., *Parametric optimization: singularities, pathfollowing and jumps*, Teubner and John Wiley, Chichester (1990).
- [6] Hettich R., Parametric Optimization: Applications and Computational Methods, Preprint, University of Trier (1985).
- [7] W. Rudin, Principles of Mathematical Analysis, (third edition), McGraw-Hill, (1976).