

5. Constrained nonlinear optimization

Version: 15-11-2011

Material (*for details see*)

- **Chapter 12 in [FKS] (pp.283-320)**

A reference e.g. L.12.2 refers to the corresponding Lemma in the book [FKS]

5.1 Introduction

We consider the minimization problem

$$(P) \quad \min f(x) \quad \text{s.t. } x \in \mathcal{F}$$

where the *feasible set* is given by

$$\mathcal{F} = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, j \in J\}$$

with $J = \{1, \dots, m\}$, $f, g_j \in C^1(\mathbb{R}^n, \mathbb{R})$

Rem. We skip additional equality constraints $h_i(x) = 0$ to avoid technical difficulties.

Recall for $\bar{x} \in \mathcal{F}$ the *active index set* $J_{\bar{x}}$:

$$J_{\bar{x}} = \{j \in J \mid g_j(\bar{x}) = 0\}$$

5.2 Optimality conditions

Def. A direction $d \in \mathbb{R}^n$ is called a *strictly feasible direction*

at $\bar{x} \in \mathcal{F}$ if: $\nabla g_j(\bar{x})^T d < 0 \quad \forall j \in J_{\bar{x}} \quad (*)$

holds and a *strictly feasible descent* direction if in addition

to $(*)$: $\nabla f(\bar{x})^T d < 0$

Th.12.4 [Necessary optimality conditions] Let $\bar{x} \in \mathcal{F}$ be a local minimizer of (P). Then the following equivalent conditions hold:

- (a) (*Primal condition*) There does not exist a strictly feasible descent direction at \bar{x} .
- (b) (*Dual condition*) There exists a non-trivial solution $\mu_0 \geq 0, \mu_j \geq 0, j \in J_{\bar{x}}$ (i.e., not all μ_j 's are = 0) of:

$$\mu_0 \nabla f(\bar{x}) + \sum_{j \in J_{\bar{x}}} \mu_j \nabla g_j(\bar{x}) = 0 \quad (\text{John condition})$$

We say: \bar{x} is a John point.

Under suitable conditions the (stronger) KKT condition (i.e. the J-cond. with $\mu_0 = 1$) will be necessary for optimality.

Def. We say that the *Mangasarian Fromovitz constraint qualification* (MFCQ) holds at $\bar{x} \in \mathcal{F}$, if there exists \hat{d} such that (*strictly feasible direction*)

$$\nabla g_j(\bar{x})^T \hat{d} < 0, \quad j \in J_{\bar{x}}.$$

Cor.12.2 Let the MFCQ condition hold at $\bar{x} \in \mathcal{F}$. If \bar{x} is a minimizer of (P) then the KKT condition holds (\bar{x} is a KKT point).: there exist $\mu_j \geq 0$, $j \in J_{\bar{x}}$, such that

$$\nabla f(\bar{x}) + \sum_{j \in J_{\bar{x}}} \mu_j \nabla g_j(\bar{x}) = 0 \quad (\text{KKT-condition})$$

Rem. Different from the convex case, in general nonlinear optimization, the KKT condition is **not sufficient** for minimality.

Take as example:

$$\min x^3 \quad \text{s.t.} \quad x \leq 0 \quad \text{with KKT-point } \bar{x} = 0 \text{ (no minimizer).}$$

Ex.12.n1 Consider (P) with convex $g_j \in C^1 \forall j$. Show that the following are equivalent:

- (a) There exists a point $\bar{x} \in \mathcal{F}$ satisfying MFCQ.
- (b) The Slater condition holds, i.e., there exists x^* such that $g_j(x^*) < 0 \quad \forall j \in J$.
- (c) At all points $\bar{x} \in \mathcal{F}$, MFCQ holds.

Def. We say that the *Linear independency constraint qualification* (LICQ) holds at $\bar{x} \in \mathcal{F}$, if the active gradients

$$\nabla g_j(\bar{x}), \quad j \in J_{\bar{x}} \quad \text{are linearly independent.}$$

Ex.12.11 Show that LICQ implies MFCQ.

First order sufficient conditions

Under additional assumptions the KKT condition can be sufficient for minimality.

Def. A point $\bar{x} \in \mathcal{F}$ is called a **strict local minimizer of order $p = 1$ or $p = 2$** if with some $c > 0, \varepsilon > 0$:

$$f(x) - f(\bar{x}) \geq c \|x - \bar{x}\|^p \quad \forall x \in \mathcal{F}, \|x - \bar{x}\| < \varepsilon$$

Th.12.2' (*first order sufficient condition*)

Let the KKT condition hold for $\bar{x} \in \mathcal{F}$ with $\mu_j > 0, j \in J_{\bar{x}}$, and

$$|J_{\bar{x}}| = n, \quad \nabla g_j(\bar{x}), j \in J_{\bar{x}}, \quad \text{linearly independent.}$$

Then \bar{x} is a strict local minimizer of (P) of order $p = 1$.

Second order sufficient optimality conditions

In general for sufficient optimality conditions also second order information is needed.

Def. Let $\bar{x} \in \mathcal{F}$. The set (of weakly feasible descent directions)

$$C(\bar{x}) := \{d \mid \nabla f(\bar{x})^T d \leq 0, \nabla g_j(\bar{x})^T d \leq 0, j \in J_{\bar{x}}\}$$

is called the cone of *critical directions* at \bar{x} .

Th.12.6 Let the KKT condition hold for $\bar{x} \in \mathcal{F}$ with multipliers $\mu_j \geq 0, j \in J_{\bar{x}}$, such that

$$d^T \left[\nabla^2 f(\bar{x}) + \sum_{j \in J_{\bar{x}}} \mu_j \nabla^2 g_j(\bar{x}) \right] d > 0 \quad \forall d \in C(\bar{x}) \setminus \{0\}.$$

Then \bar{x} is a strict local minimizer of (P) of order $p = 2$.

5.3 Feasible descent method (primal method)

Basic idea. *This method is a generalization of the “steepest descent method” to constrained programs; it is based on “feasible descent directions”.*

Consider again

$$(P) \quad \min_{x \in \mathcal{F}} f(x), \quad \mathcal{F} = \{x \mid g_j(x) \leq 0, j \in J\}$$

Recall: d_k is a strictly feasible descent dir. in $x_k \in \mathcal{F}$ if

$$\nabla f(x_k)^T d_k < 0, \quad \nabla g_j(x_k)^T d_k < 0, \quad \forall j \in J_{x_k}$$

Ex. Let d_k be a strictly feasible descent direction in $x_k \in \mathcal{F}$. Show that for any $t > 0$, small enough:

$$f(x_k + td_k) < f(x_k) \quad \text{and} \quad x_k + td_k \in \mathcal{F}$$

Recall Theorem 12.4. There is no strictly feasible descent direction d_k at x_k iff x_k is a John point.

This observation leads to the following algorithm.

Feasible Direction Method (*Zoutendijk*)

INIT: Choose a starting point $x_0 \in \mathcal{F}$

ITER: WHILE x_k is not a John point DO

BEGIN

 Choose a strictly feasible descent direction d_k

 Determine a solution t_k for the problem

$$\min_{t>0} \{f(x_k + td_k) \mid x_k + td_k \in \mathcal{F}\}$$

 Set $x_{k+1} = x_k + t_k d_k$.

END

How to compute d_k ?

Naive method: Solve the linear program ,

$$(\star) \quad \begin{array}{ll} \min_{d,z} z & \text{s.t.} \quad \nabla f(x_k)^T d - z \leq 0 \\ & \nabla g_j(x_k)^T d - z \leq 0 \quad j \in J_{x_k} \\ & \pm d_i \leq 1 \quad \forall i \end{array}$$

where $\pm d_i \leq 1$ is added to guarantee a finite optimal solution.

Remark. If the sol. (z, d) in (\star) satisfies $z < 0$ then d is a strictly feasible descent direction and vice versa.

Remark. This “naive method” may result in *jamming* in the index sets J_{x_k} , leading to “bad” convergence.

Possibly $x_k \rightarrow \bar{x}$ where \bar{x} is *not* a John point (see Wolfe’s example [FKS, p292].)

Better: The following method of Topkis and Veinott takes *all* constraints into account: compute d_k by solving

$$LP_{x_k} : \quad \min_{d,z} z \quad \text{s.t.} \quad \begin{aligned} \nabla f(x_k)^T d - z &\leq 0 \\ \nabla g_j(x_k)^T d - z &\leq -g_j(x_k) \quad \forall j \in J \\ \pm d_i &\leq 1 \quad \forall i \end{aligned}$$

Ex. LP_{x_k} has optimal value $z = 0$ precisely when x_k satisfies the John optimality conditions

This method leads to a “positive” convergence result.

Th.12.5 [Topkis and Veinott]

Assume that the Feasible Direction Method with d_k computed by LP_{x_k} generates the points x_ℓ . Suppose $x_S \rightarrow \bar{x}$ for an infinite subset $S \subset \mathbb{N}$. Then \bar{x} is a John point.

Convergence properties: *Roughly speaking, these feasible descent methods which only use first order (gradient) information show typically a (often slow) linear convergence similar to the steepest descent method in unconstrained optimization.*

5.4 Penalty methods

Basic Idea: Transform the constrained problem

$$(P) \quad \min_{x \in \mathcal{F}} f(x), \quad \mathcal{F} = \{x \mid g_j(x) \leq 0, j \in J\}$$

into an unconstrained program.

Penalty method: Instead of (P) solve the unconstrained problem (with suitable parameter $r > 0$; suitable $p(x)$):

$$(P_r) \quad \min_{x \in \mathbb{R}^n} p_r(x) = f(x) + r p(x)$$

Assumption: The feasible set \mathcal{F} is non-empty.

Def. $p : \mathbb{R}^n \rightarrow \mathbb{R}$ is a *penalty function* with respect to \mathcal{F} if

$$p(x) = 0 \quad \forall x \in \mathcal{F} \quad \text{and} \quad p(x) > 0 \quad \forall x \in \mathbb{R}^n \setminus \mathcal{F}$$

Examples. Setting $g_j^+(x) = \max\{0, g_j(x)\}$:

$$p(x) = \sum_{j \in J} g_j^+(x) \quad \text{or} \quad p(x) = \sum_{j \in J} (g_j^+(x))^2.$$

are such penalty functions.

Note. Typically $g_j^+ \notin C^1$ at \bar{x} with $g_j(\bar{x}) = 0$.

Ex.12.21 Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be a C^1 -function. Then also $\hat{g}(x) = (g^+(x))^2$ is a C^1 -function. Determine $\nabla \hat{g}(x)$.

Th.12.9 [general convergence result]

Let $p(x)$ be a continuous penalty function wrt. $\mathcal{F} \neq \emptyset$. Choose parameters $r_k \rightarrow \infty$ and consider global minimizers x_k of the problems (P_{r_k}) . If $x_k \rightarrow x^*$ then x^* is a global minimizer of (P).

5.4.1 Exact penalty function

We consider the “*exact*” penalty problem (non-smooth)

$$(P_r) \quad \min_{x \in \mathbb{R}^n} p_r(x) = f(x) + r \sum_{j \in J} g_j^+(x)$$

The next theorem explains why this method is called “*exact*”.

Th. 12.10. Let (LICQ) be satisfied for the local minimizer \bar{x} of (P) and assume that the sufficient optimality conditions of Th.12.6 are fulfilled with multipliers $\bar{\mu}_j > 0, \forall j \in J_{\bar{x}}$. Then \bar{x} is also a (“*exact*”) local minimizer of the penalty problem (P_r) whenever $r > \max\{\bar{\mu}_j \mid j \in J_{\bar{x}}\}$.

Ex. (to illustrate the working of the different penalty functions)

Solve the program: $\min x^2$ s.t. $g(x) := 1 - x \leq 0$

by the penalty method with $p(x) = (g^+(x))^2$ and by the exact penalty method with $p(x) = g^+(x)$.

The proof of Th.12.10 is based on the following “perturbation arguments”.

observation: Under LICQ, a local minimizer \bar{x} of (P) can be computed as a solution of (*necessary KKT conditions*):

$$(*) \quad \begin{aligned} \nabla f(x) + \sum_{j \in J_{\bar{x}}} \mu_j \nabla g_j(x) &= 0 \\ g_j(x) &= 0, \quad j \in J_{\bar{x}} \end{aligned}$$

For small perturbation vectors $0 \approx u \in \mathbb{R}^m$ we consider the *perturbed program*

$$(P_u) \quad \min_x f(x), \quad \text{s.t.} \quad g_j(x) \leq u_j, \quad j \in J$$

with minimizer $x(u)$, minimal value $v(u)$ and perturbed KKT system:

$$\begin{aligned} \nabla f(x) + \sum_{j \in J_{\bar{x}}} \mu_j \nabla g_j(x) &= 0 \\ g_j(x) &= u_j, \quad j \in J_{\bar{x}} \end{aligned}$$

depending on the parameter u

By applying the Inverse Function Theorem (under some regularity conditions) it can be shown that (locally around $\bar{u} = 0$) the functions $\mathbf{x}(\mathbf{u})$ and $\mathbf{v}(\mathbf{u})$ are \mathbf{C}^1 -functions of \mathbf{u} .

Th.12.8 Assume that (LICQ) is satisfied at the KKT-point $\bar{\mathbf{x}} \in \mathcal{F}$ and that the sufficient optimality conditions of Th.12.6 hold with multipliers $\bar{\mu}_j > 0, \forall j \in \mathbf{J}_{\bar{\mathbf{x}}}$. Then there exists a neighborhood U of $\bar{\mathbf{u}} = 0$ and a continuously differentiable function $\mathbf{x} : U \rightarrow \mathbb{R}^n$ such that $\mathbf{x}(0) = \bar{\mathbf{x}}$ and for all $\mathbf{u} \in U$,

- (i) $\mathbf{x}(\mathbf{u})$ is a strict local minimizer for $(P_{\mathbf{u}})$.
- (ii) $\mathbf{v}(\mathbf{u}) = f(\mathbf{x}(\mathbf{u}))$ satisfies $\nabla \mathbf{v}(0) = -\bar{\boldsymbol{\mu}}$

5.5 SQP-method, Sequentially quadratic programming

This method leads to algorithms for solving constrained programs that similarly to the Quasi-Newton methods in unconstrained optimization are superlinearly convergent.

Idea: To solve

$$(P) \quad \min_{x \in \mathcal{F}} f(x), \quad \mathcal{F} = \{x \mid g_j(x) \leq 0, j \in J\}$$

better than in the *primal* method we try to compute a “feasible descent direction” d_k by solving a suitable *quadratic* subproblem (also using second order information).

To find d_k , for given (x_k, μ_k) , possibly $x_k \notin \mathcal{F}$, we solve the quadratic program

$$(Q_k) \quad \min_d \quad \nabla f(x_k)^T d + \frac{1}{2} d^T L_k d \quad \text{s.t.} \\ \nabla g_j(x_k)^T d + g_j(x_k) \leq 0 \quad \forall j \in J$$

where $L_k = \nabla_x^2 L(x_k, \mu_k)$ (or, an approximation thereof).

Rem. Note that (Q_k) takes *all* inequality constraints into account, similarly to the primal method of Topkis-Veinott.

Computing a minimizer d_k of (Q_k) we have to consider the cases $d_k = 0$ and $d_k \neq 0$.

Lem.12.4 If $d_k = 0$ is a minimizer of (Q_k) , with corresponding multiplier μ_{k+1} then x_k is a KKT-point for (P) with multiplier μ_{k+1} .

If $d_k \neq 0$ is the minimizer of (Q_k) , we use d_k as search direction for the next iteration:

$$x_{k+1} = x_k + t d_k .$$

A problem appears: $d_k \neq 0$ need not be a descent direction for $f(x)$ (see [FKS, Ex.12.25]).

Ex. If $x_k \in \mathcal{F}$ and $d_k \neq 0$ is a solution of (Q_k) with $L_k \succ$ (pd.) then $\nabla^T f(x_k) d_k < 0$, i.e., d_k is a descent direction for f .

However $d_k \neq 0$ is always a descent direction for the exact penalty function (for suitable $r > 0$)

$$p_r(x) = f(x) + r \sum_{j \in J} g_j^+(x)$$

In this context, p_r is called a *merit function*

Th.12.12 Let the matrix L_k be positive definite and let $d_k \neq 0$ a minimizer for (Q_k) with corresponding multiplier μ_{k+1} . Then d_k is a descent direction for p_r , i.e.,

$$p_r(x_k + td_k) < p_r(x_k) \quad \text{for small } t > 0 ,$$

provided $r \geq \max\{(\mu_{k+1})_j \mid j \in J\}$.

Under the hypothesis of Th.12.10, a local minimizer of (\mathbf{P}) is also a local minimizer of the merit function \mathbf{p}_r . So Th.12.12 suggests the following algorithm for computing a local minimizer of $\mathbf{p}_r(\mathbf{x})$ and hence a candidate minimizer of (\mathbf{P}) .

SQP-Method with merit function \mathbf{p}_r , (possibly $x_k \notin \mathcal{F}$)

INIT: Choose $\mathbf{x}_0 \in \mathbb{R}^n$, $r > 0$ large enough,

$L_0 \in \mathbb{R}^{n \times n}$ positive definite

ITER: WHILE \mathbf{x}_k is not a KKT-point DO

BEGIN

Compute a solution \mathbf{d}_k of subproblem (Q_k) ,

Compute a solution t_k for the problem

$$(*) \quad \min_{t > 0} p_r(\mathbf{x}_k + t\mathbf{d}_k)$$

Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k \mathbf{d}_k$ and choose L_{k+1} positive definite.

END

REMARK. Concerning the convergence properties of the SQP-method.

- For equality constrained programs, *i.e.*, $g_j(x) = 0$, the solution d_k of (Q_k) with $L_k := \nabla_x^2 L(x_k, \mu_k)$, coincides with the Newton iteration step at x_k, μ_k applied to the KKT system (*, p.16). So, under suitable assumptions the SQP method with (*full step*) $x_{k+1} = x_k + d_k$ is locally quadratically convergent to a local minimizer \bar{x} of (P) (see [FKS, Ex.12.26]).
- Under suitable assumptions, if the SQP method generates t_k, d_k and $x_{k+1} = x_k + t_k d_k$ such that

$$t_k \rightarrow 1 \quad \text{and} \quad x_k \rightarrow \bar{x},$$

then superlinear convergence occurs.

- **[The Maratos Effect:]** Unfortunately, $t_k \rightarrow 1$ is not generally true for the line minimization (*) of the SQP-algorithm. It may even happen for x_k arbitrarily close to the solution \bar{x} of (P) that the full Newton step $x_{k+1} = x_k + d_k$ increases the merit function to

$$p_r(x_k + d_k) > p_r(x_k)$$

(see [FKS, Ex.12.25] for an illustrative example). This negative effect, destroying the superlinear convergence, was discovered by Maratos. Fortunately, it is rarely observed in practice.