On sequential optimality conditions for constrained optimization

José Mario Martínez
www.ime.unicamp.br/∼martinez

UNICAMP, Brazil

2011

Collaborators

This talk is based in joint papers with Roberto Andreani (Campinas), Benar F. Svaiter (IMPA), Lucio T. Santos (Campinas) and Gabriel Haeser (U. F. São Paulo).

1. R. Andreani, J. M. Martínez and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization* 20, pp. 3533-3554 (2010).
2. R. Andreani, G. Haeser, J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. To appear in *Optimization* 2011.
3. J. M. Martínez and B. F. Svaiter. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications* 118, pp. 117-133 (2003).

Nonlinear Programming Problem

NLP

Minimize $f(x)$

subject to

$$h(x) = 0, g(x) \leq 0,$$

where $x \in \mathbb{R}^n$, $h(x) \in \mathbb{R}^m, g(x) \in \mathbb{R}^p$ and all the functions are smooth.

- Finding global minimizers is difficult.
- Affordable algorithms for solving large-scale problems usually guarantee convergence only to “stationary points”.
- Affordable algorithms are useful tools in the process of finding global solutions (Multistart and so on).

Pointwise Optimality conditions

- Stationary points are points that satisfy some necessary optimality condition.
- In Nonlinear Programming, (pointwise) Necessary Optimality conditions take the form:
  The point x^* is feasible and fulfills the KKT conditions OR fails to satisfy the XXX constraint qualification.
- The strength of a pointwise necessary optimality condition is associated with the weakness of the constraint qualification. (Weak constraint qualifications are good.)

Sequential Optimality conditions

Nonlinear Programming algorithms are iterative. They generate sequences $\{x^k\}$ that, presumably, converge to stationary points which, in fact, are never reached exactly.
So, any practical algorithm needs to decide, at each iterate x^k whether x^k is an approximate solution (approximate stationary point) or not. Stopping criterion!.
This motivates to study Sequential Optimality conditions.

- A feasible point x^* satisfies a Sequential Optimality condition if there exists a sequence that converges to x^* and fulfills some property PPP.
- Local minimizers satisfy Sequential Optimality conditions.
- The fulfillment of PPP may be verified at each iterate of a practical algorithm.

Example: Approximate Gradient Projection AGP

A feasible point x^* satisfies AGP (Martínez-Svaiter 2003) if there exists a sequence $\{x^k\} \to x^*$ such that the projected gradient $P(x^k - \nabla f(x^k)) - x^k$ on a linear approximation of the constraints tends to zero.

AGP is a strong optimality condition: Every local minimizer satisfies AGP and AGP implies the pointwise optimality condition KKT or not-CPLD.

(CPLD is a weak constraint qualification that says that positively linear dependent gradients of active constraints at x^* remain linearly dependent in a neighborhood of x^*).

AGP generates the natural stopping criterion for "Inexact Restoration Methods" and other algorithms for Nonlinear Programming.

Approximate KKT condition

We say that a feasible point x^* satisfies AKKT if there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p$ such that

$$\lim_{k\to\infty} x^k = x^*,$$

$$\lim_{k\to\infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0$$

and

$$\lim_{k\to\infty} \min\{-g_i(x^k), \mu_i^k\} = 0, \quad \text{for all} \quad i = 1, \dots, p.$$

Recall that a feasible point x^* satisfies KKT if there exist $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p$ such that

$$\nabla f(x^*) + \nabla h(x^*)\lambda + \nabla g(x^*)\mu = 0$$

and

$$\min\{-g_i(x^*), \mu_i\} = 0, \quad \text{for all} \quad i = 1, \ldots, p.$$

Consequence: KKT implies AKKT. (Take $x^k = x^*, \lambda^k = \lambda, \mu^k = \mu$ for all k).

Properties of the Approximate KKT condition

- Every local minimizer satisfies AKKT (Even if it does not satisfy KKT). (No constraint qualification is needed.)
- AKKT is a strong optimality condition. (It implies KKT or not-CPLD.)
- AKKT generates the natural stopping criterion

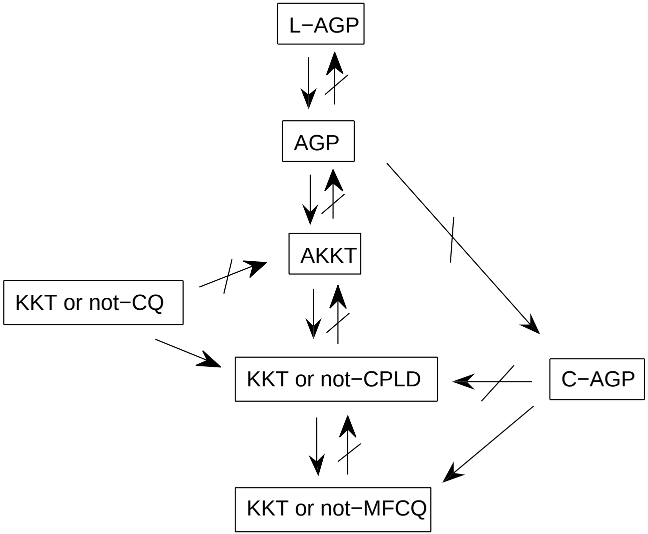
$$\|h(x^k)\| \leq \varepsilon, \|g(x^k)_+\| \leq \varepsilon,$$

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| \leq \varepsilon$$

and

$$|\min\{-g_i(x^k), \mu_i^k\}| \leq \varepsilon \quad \text{for all} \quad i = 1, \ldots, p.$$

- Sequences generated by well-established Nonlinear Programming algorithms satisfy the Approximate KKT condition.


L−AGP
AGP
AKKT
KKT or not−CQ
KKT or not−CPLD
C−AGP
KKT or not−MFCQ

Complementary Approximate KKT condition (CAKKT)

CAKKT is a new sequential optimality condition (Andreani, Martínez, Svaiter, 2010).
We say that a feasible point x^* satisfies AKKT if there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\lambda^k\} \subset \mathbb{R}^m$, $\{\mu^k\} \subset \mathbb{R}^p_+$ such that

$$\lim_{k\to\infty} x^k = x^*,$$

$$\lim_{k\to\infty} \|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| = 0,$$

$$\lim_{k\to\infty} g_i(x^k)\mu_i^k = 0, \quad \text{for all} \quad i = 1, \ldots, p$$

and

$$\lim_{k\to\infty} h_i(x^k)\lambda_i^k = 0, \quad \text{for all} \quad i = 1, \ldots, m.$$

Properties of the Complementary Approximate KKT condition

- Every local minimizer x^* satisfies CAKKT. (Even if x^* is not KKT.) (Independently of constraint qualifications.)
- CAKKT is strictly stronger than AKKT and strictly stronger than the Approximate Gradient Projection (AGP) Condition.
- CAKKT is strictly stronger than the pointwise optimality condition KKT or not-CPLD.

- CAKKT generates the natural? stopping criterion

$$\|h(x^k)\| \leq \varepsilon, \|g(x^k)_+\| \leq \varepsilon,$$

$$\|\nabla f(x^k) + \nabla h(x^k)\lambda^k + \nabla g(x^k)\mu^k\| \leq \varepsilon,$$

$$|g_i(x^k)\mu_i^k| \leq \varepsilon \text{ for all } \quad i = 1, \dots, p$$

and

$$|h_i(x^k)\lambda_i^k| \leq \varepsilon \text{ for all } \quad i = 1, \dots, m.$$

- Sequences generated by well-established Nonlinear Programming algorithms satisfy CAKKT Really?.

Example: Augmented Lagrangian (Algencan)

At each outer iteration we compute x^k, an approximate stationary point of the augmented Lagrangian

$$L(x, \lambda^k, \mu^k) = f(x) + \frac{\rho_k}{2}\left[\left(h(x) + \frac{\lambda^k}{\rho_k}\right)^2 + \left(g(x) + \frac{\mu^k}{\rho_k}\right)_+^2\right].$$

We update

$$\lambda^{k+1} = \lambda^k + \rho_k h(x^k) \text{ and } \mu^{k+1} = (\mu^k + \rho_k g(x^k))_+.$$

We update $\rho_{k+1} = \gamma\rho_{k+1}$ $(\gamma > 1)$ if $\max\{\|h(x^k\|, \|\max(g(x^k), \mu^{k+1})\| > r\max\{\|h(x^{k-1}\|, \|\max(g(x^{k-1}), \mu^k)\|$. Else, we maintain $\rho_{k+1} = \rho_k$.

We project λ^{k+1} and μ^{k+1} on safeguarding boxes.

Generalized Lojasiewicz Inequality (GLI)

$f : \mathbb{R}^n \to \mathbb{R}$ satisfies GLI at x if

$$|f(z) - f(x)| \leq \varphi(z)\|\nabla f(z)\|$$

for all z in a neighborhood of x, where φ is continuous and $\varphi(x) = 0$

Every reasonable (for example, analytic) function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies GLI.

We will assume that the constraint functions h_i, g_j of our NLP problem satisfy GLI at the feasible points

Augmented Lagrangian Algorithm and CAKKT

Theorem

Assume that x^* is a feasible limit point of a sequence generated by the Augmented Lagrangian method and that the GLI assumption holds at x^*. Then, the associated subsequence $(x^k, \lambda^{k+1}, \mu^{k+1})$ fulfills the Complementary Approximate KKT condition.

Consequence: At least one reasonable NLP algorithm converges to CAKKT points.

GLI Assumption cannot be eliminated

Sequences generated by the Augmented Lagrangian method generate Approximate KKT sequences without the GLI condition. However, for convergence to Complementary Approximate KKT points, the GLI condition cannot be eliminated.

Example:

Consider the problem

$$\text{Minimize } x \ \text{ subject to } \ h(x) = 0,$$

where

$$h(x) = x^4 \sin\left(\frac{1}{x}\right) \ \text{ if } \ x \neq 0,$$

and $h(0) = 0$. h does not fulfill GLI at $x = 0$.

We are able to define an instance of the algorithm in which the sequence x^k converges to $x^* = 0$ and the condition $h(x^k)\lambda^{k+1} \to 0$ does not hold.

What happens with other reasonable NLP algorithms?

Essentially, we don't know.
We performed a good number of numerical experiments with SQP (Sequential Quadratic Programming) in cases in which the solution of the problem is not KKT.
We observed that the approximate complementarity condition $\lambda^k h(x^k) \to 0$ generally holds but the Lagrangian condition $\nabla f(x^k) + \nabla h(x^k)\lambda^k \to 0$ does not!
So, contrary to expected, SQP seems to fail to satisfy CAKKT not because of the "C" but because of the "KKT".

Should CAKKT be used as stopping criterion?

(Perhaps Naive) Answer:
If one is using an algorithm that provably generates CAKKT sequences, CAKKT should be used as stopping criterion since, eventually, it will be satisfied.
If your algorithm does not generate CAKKT sequences, it seems better not to use CAKKT at all.