Degenerate Nonlinear Programs

Mihai Anitescu

Unconstrained Optimization

$$\min f(x)$$

At points x^* at which the quadratic growth (QG) condition holds

$$f(x) \geq f(x^*) + \sigma ||x \Leftrightarrow x^*||^2 \quad x \in B(x^*, r)$$

- Steepest descent: $f(x) \to f(x^*)$ Q-linearly.
- Newton method $x \to x^*$ Q-linearly.

Constrained Optimization

$$\min_{x \in D} f(x)$$

Do the same good algorithmic properties hold when feasible quadratic growth is satisfied?

$$f(x) \geq f(x^*) + \sigma ||x \Leftrightarrow x^*||^2, \quad \forall x \in D \cap B(x^*, r)$$

Motivation: The study of convergence properties under very general conditions may result in more robust algorithms for large-scale programming. Robustness: The ability of maintaining a good local rate of convergence when the traditional analysis assumptions are only marginally satisfied.

Rates of Convergence

- $x^k \to x^*$ R-linearly if $\limsup \sqrt[k]{||x^k \Leftrightarrow x^*||} \to c < 1$.
- $x^k \to x^*$ Q-linearly if $\limsup \frac{||x^{k+1} - x^*||}{||x^k - x^*||} \to c < 1$.
- $x^k \to x^*$ superlinearly if $\limsup \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = 0$.

Nonlinear Program (NLP)

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & h_j(x) = 0 & i = 1 \ldots r \\ & g_j(x) \le 0 & j = 1, \ldots m \end{array}$$

$x \in \mathbb{R}^n$, f, g, h are sufficiently smooth.

KKT conditions

The Lagrangian:

$$\begin{aligned}\mathcal{L}(x,\mu,\lambda) &= f(x) + \sum_{i=1}^{m} \mu_i h_i(x) + \sum_{i=1}^{r} \lambda_j g_j(x) \\ &= f(x) + \mu^T h(x) + \lambda^T g(x)\end{aligned}$$

Stationary point of NLP : A point x for which there exist $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^r$ such that

$$\nabla_x \mathcal{L}(x,\lambda,\mu) = 0, \quad h(x) = 0, \quad g(z) \leq 0, \quad (\lambda)^T g(z) = 0$$

KKT theorem: under certain constraint qualification conditions, the solution x^* of the NLP is a stationary point of the NLP.

The active set of a feasible $x \in \mathbb{R}^n$:

$$\mathcal{A}(x) = \{j | 1 \leq j \leq m,\ g_j(x) = 0\}$$

Steepest Descent Direction for an NLP

Unconstrained Optimization:

$$d = \Leftrightarrow \nabla f(x) = \arg\min\{\frac{1}{2}d^T d + \nabla f(x)^T d\}$$

Constrained Optimization: d is the solution of the Quadratic Program (QP) with linearized constraints:

$$\begin{array}{lll} \text{minimize} & \nabla f(x)^T d + \frac{1}{2} d^T d & \\ \text{subject to} & h_i(x) + \nabla h_i(x)^T d = 0 & i = 1, \ldots, r \\ & g_j(x) + \nabla g_j(x)^T d \leq 0, & j = 1, \ldots, m. \end{array}$$

The QP is feasible whenever x is feasible, regardless of the satisfiability of first-order conditions. d is unique (if QP is feasible) and $d = 0$ iff x is a stationary point of the NLP.

Robinson's Example

$$\begin{array}{rl} \min & f(x) = \frac{x^2}{2} \\ \text{subject to} & h(x) = x^6 \sin \frac{1}{x} \end{array}$$

- $x = \frac{1}{\pm k\pi}$, $k \in \mathbb{N}$, $k \neq 0$ are stationary points accumulating to zero.
- The direction of steepest descent $d = 0$. Thus QG alone will not induce $x^k \to x^* = 0$, even when started arbitrarily close to x^*.
- The feasible set needs to satisfy a constraint qualification.
- For steepest descent, the issue of isolated stationary points is fundamental.

Traditional Constraint Qualifications (KKT holds)

- Linear Independence CQ (LICQ):

$$\nabla h_i(x^*),\ i = 1, \dots, r \quad \text{and} \nabla g_j(x^*), j \in \mathcal{A}(x^*)$$

are linearly independent. λ^* satisfying KKT is unique.

- Linear Constraint CQ: $h(x)$ and $g(x)$ are linear.
- NLP not satisfying LICQ are called degenerate.

Mangasarian-Fromowitz Constraint Qualification

- Mangasarian Fromowitz CQ (MFCQ): $\nabla h_j(x^*), 1 \leq j \leq r$ are linearly independent and

$$\exists p \in R^n \quad \text{such that} \quad \begin{aligned} &\nabla_x h_j(x^*)^T p = 0,\ j = 1, \ldots m \\ &\nabla_x g_i(x^*)^T p < 0,\ i \in \mathcal{A}(x^*). \end{aligned}$$

- MFCQ holds $\Leftrightarrow$ The set $\mathcal{M}(x^*)$ of the multipliers satisfying KKT is bounded.

- The critical cone:

$$\begin{aligned} \mathcal{C} \;=\; &\{u \in \mathbb{R}^n | \nabla h_i(x^*)^T u = 0,\ 1 \leq i \leq r, \\ &\nabla g_i(x^*)^T u \leq 0,\ i \in \mathcal{A}(x^*),\ \nabla f(x)^T u = 0\} \end{aligned}$$

Second-Order Sufficient Conditions

- Traditional SOSC (second-Order Sufficient Conditions) that x^* be a strict local minimum: LICQ and

$$u^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) u > 0, \forall u \in \mathcal{C}.$$

- Relaxed SOSC (in Fiacco): MFCQ and

$$\exists (\mu^*, \lambda^*) \in \mathcal{M}(x^*), \text{ such that } u^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) u > 0, \ \forall u \in \mathcal{C}.$$

- Shapiro SOSC: MFCQ and

$$\begin{aligned} &\forall u \in \mathcal{C}, \ \exists (\mu^*, \lambda^*) \in \mathcal{M}(x^*), \text{such that} \\ &u^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) u > 0. \end{aligned}$$

- Shapiro SOSC $\Leftrightarrow$ Quadratic Growth and MFCQ !

SOSC for isolated stationary points

- Traditional SOSC ensures it via implicit function theorem, if $\lambda^*_{\mathcal{A}(x^*)} > 0$ (strict complementarity).
- Robinson SOSC: MFCQ and

$$\forall(\mu^*, \lambda^*) \in \mathcal{M}(x^*),\ \forall u \in \mathcal{C} \quad u^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) u > 0.$$

  The L_∞ penalty function using the steepest descent direction induces Q-linear convergence to x^* (M).
- Quadratic growth + MFCQ $\Rightarrow$ x^* is an isolated stationary point (M)! The steepest descent direction will not be zero in $B(x^*, r)$. Thus linear convergence may be achievable even in these very general conditions.

Superlinear Convergence for Traditional SOSC

Assume $\mathcal{A}(x^*) = \{1, \ldots, m\}$. LICQ and strict complementarity ensure that the Newton step for the KKT

$$\nabla L(x, \lambda) = 0, \quad g(x) = 0$$

is well defined near x^*, λ^*.

$$\begin{pmatrix} \nabla_{xx}\mathcal{L}(x^k, \lambda^k) & \nabla g(x^k) \\ \nabla g(x^k) & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = \begin{pmatrix} \Leftrightarrow \nabla_x \mathcal{L}(x^k, \lambda^k) \\ \Leftrightarrow g(x^k) \end{pmatrix} =$$

$$x^{k+1} = x^k + \Delta x, \quad \lambda^{k+1} = \lambda^k + \Delta \lambda$$

Then $(x^k \lambda^k) \to (x^*, \lambda^*)$ quadratically.

Superlinear Convergence for Relaxed SOSC

- Starting with (x, λ) near (x^*, λ^*) that satisfies the Relaxed SOSC and strict complementarity. Then the stabilized Newton method

$$\begin{pmatrix} \nabla_{xx}\mathcal{L}(x^k, \lambda^k) & \nabla g(x^k) \\ \nabla g(x^k) & \Leftrightarrow \mu^k \end{pmatrix} \begin{pmatrix} x^{k+1} \Leftrightarrow x^k \\ \lambda^{k+1} \Leftrightarrow \lambda^k \end{pmatrix} = \begin{pmatrix} \Leftrightarrow \nabla_x \mathcal{L}(x^k, \lambda^k) \\ \Leftrightarrow g(x^k) \end{pmatrix} =$$

- Then $(x^k, \lambda^k) \to (x^*, \lambda^*)$ superlinearly if $\mu^k = \Omega||(x^k, \lambda^k) \Leftrightarrow (x^*, \lambda^*)||$.
- By Schur Complement and since there exists a positive definite augmented Lagrangian, the system is nonsingular.
- The method has been extended to cases without strict complementarity, for stronger second-order conditions of the Lagrangian.

The L_∞ exact penalty function

- For simplicity, only inequality constraints will be considered.
- Need a measure that will balance feasibility and optimality (see Sven's Filter SQP). This will measure progress along a given direction.
- The L_∞ penalty function

$$P(x) = \max\{g_0(x), g_1(x), ...g_m(x)\}.$$

Here $g_0(x) \equiv 0$.

- x^* is an unconstrained minimum of the penalized objective function $\phi(x) = f(x) + c_\phi P(x)$.
- However, $\phi(x)$ becomes nondifferentiable.

Descent Directions for $\phi(x)$

$$\begin{array}{ll} \text{minimize} & \nabla f(x)^T d + \frac{1}{2} d^T H d + c_\phi \zeta \\ \text{subject to} & g_j(x) + \nabla g_j(x)^T d \le \zeta, \qquad j = 0, 1, 2 \ldots m, \end{array}$$

- If λ is a multiplier, $c_\phi = \lambda_0 + \sum_{i=1}^m \lambda_i$ and $\lambda_0 \zeta = 0$ (this QP is always feasible).
- If $H = I$ and

$$c_\phi > 2\gamma + \sum_{i=1}^m \lambda_i^*, \quad \forall \lambda^* \in \mathcal{M}(x^*)$$

then $\zeta = 0$ and d is the steepest descent direction. With MFCQ $\mathcal{M}(x^*)$ is bounded.

- With MFCQ, the feasible set has an interior and the steepest descent QP is always feasible.

L_∞ SQP algorithm near x^*

SQP: Sequential Quadratic Programming.

1. Set $k = 0$, choose x^0.
2. Compute d^k from

$$\begin{array}{ll} \text{minimize} & \nabla f(x^k)^T d + \frac{1}{2} d^T d \\ & g_j(x^k) + \nabla g_j(x^k)^T d \leq 0, \qquad j = 1, \ldots, m. \end{array}$$

3. Choose α^k from a line search procedure, and set $x^{(k+1)} = x^k + \alpha^k d^k$.
4. Set $k = k + 1$ and return to Step 2.

Step size selection

(a) Minimization rule Here α^k is chosen such that

$$\phi(x^k + \alpha^k d^k) = \min_{\alpha \geq 0} \{\phi(x^k + \alpha d^k).\}$$

(b) Limited minimization rule Here a fixed scalar $s > 0$ is selected, and α^k is chosen such that

$$\phi(x^k + \alpha^k d^k) = \min_{\alpha \in [0,\, s]} \{\phi(x^k + \alpha d^k)\}.$$

(c) Armijo rule Here fixed scalars s, τ, and σ with $s > 0$, $\tau \in (0, 1)$, and $\sigma \in (0, \frac{1}{2})$ are chosen and we set $\alpha^k = \tau^{m_k} s$, where m_k is the first nonnegative integer m for which

$$\phi(x^k) \Leftrightarrow \phi(x^k + \tau^m s d^k) \geq \sigma \tau^m s (d^k)^T d^k.$$

It can be shown that the Armijo rule yields a stepsize after a finite number of iterations.

Main Theorem

If x^* satisfies MFCQ and the Quadratic Growth Condition

$$f(x) \geq f(x^*) + \sigma ||x \Leftrightarrow x^*||^2, \quad \forall x \text{ feasible in } B(x^*, r)$$

If x^0 is sufficiently close to x^*, with x^k generated by the steepest descent algorithm with an exact L_∞ penalty function with sufficiently large c_ϕ,

- $x^k \to x^*$ R-linearly.
- $\phi(x^k) \to \phi(x^*)$ Q-linearly.
- x^* is an isolated stationary point of the NLP.

Shapiro SOSC $\not\Rightarrow$ Relaxed SOSC

- The example

$$
\begin{array}{lrcll}
 & \min z & & & \\
\text{sbj.to:} & g_0(x,y,z) & = & (x \Leftrightarrow 1)^2 \Leftrightarrow 2(y \Leftrightarrow 1)^2 \Leftrightarrow z & \leq 0 \\
 & g_1(x,y,z) & = & \Leftrightarrow \frac{1}{2}((x \Leftrightarrow 1)^2 + (y \Leftrightarrow 1)^2) & \\
 & & + & 3(x \Leftrightarrow 1)(y \Leftrightarrow 1) \Leftrightarrow z & \leq 0 \\
 & g_2(x,y,z) & = & \Leftrightarrow 2(x \Leftrightarrow 1)^2 + (y \Leftrightarrow 1)^2 \Leftrightarrow z & \leq 0 \\
 & g_3(x,y,z) & = & \Leftrightarrow \frac{1}{2}((x \Leftrightarrow 1)^2 + (y \Leftrightarrow 1)^2) & \\
 & & \Leftrightarrow & 3(x \Leftrightarrow 1)(y \Leftrightarrow 1) \Leftrightarrow z & \leq 0.
\end{array}
$$

- Each constraint is obtained from the other by rotating the (x, y) plane with $\frac{\pi}{4}$.

Example

- At $(1, 1, 0)$, the NLP satisfies satisfies both Quadratic Growth and MFCQ.
- However,

$$u^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) u > 0, \ \forall u \in \mathcal{C}.$$

is not satisfies by any feasible λ^*.

- For this example there will be no locally convex augmented Lagrangian ! For any $\lambda^* \in \mathcal{M}(x^*)$,

$$\nabla_{xx} \mathcal{L}(x^*, \lambda^*) + \frac{1}{\mu} \nabla g(x^*) \nabla g(x^*)^T \not\succeq 0$$

Lancelot : The Augmented Lagrangian Approach

- The feasible set is represented by

$$g_i(x) + t_i = 0,\ t_i \geq 0 \text{ for } i = 1, \ldots, m.$$

- A penalty term (with parameter μ) is added to the objective

$$\begin{array}{ll} \min & f(x) + \sum_{i=1}^{4} [\lambda_i (g_i(x) + t_i) + \frac{1}{\mu}(g_i(x) + t_i)^2] \\ \text{subject to} & t_i \geq 0, \quad i = 1, \ldots, m. \end{array}$$

- Take λ, $\mu \Rightarrow$ get $x(\lambda, \mu)$ subject to trust-region constraints $\Rightarrow$ update λ, μ.
- Desired outcome: μ bounded bellow and trust region inactive.
- In our example: Inactive trust region $\Rightarrow$ positive semidefinite augmented Lagrangian $\Rightarrow \mu \to 0$ (or otherwise would approach one of the solution augmented Lagrangians)!

The necessary conditions for Lancelot

$$\nabla_{(x,t)(x,t)} L\big|_{(x^*,0)} =$$
$$\begin{pmatrix} F_{xx} + \sum_{i=1}^4 (\lambda_i G_{xx} + \frac{2}{\mu}\nabla g_i(x^*)\nabla g_i(x^*)^T) & \frac{2}{\mu}\nabla g(x^*) \\ \frac{2}{\mu}\nabla g(x^*)^T & \frac{2}{\mu} I_4 \end{pmatrix}$$

is positive semidefinite on the subspace $t = 0$, which implies

$$0 \preceq F_{xx} + \sum_{i=1}^4 (\lambda_i G_{xx} + \frac{2}{\mu}\nabla g_i(x^*)\nabla g_i(x^*)^T) = \begin{pmatrix} \sum_{i=1}^4 \lambda_i Q_i & 0 \\ 0 & \frac{2}{\mu} \end{pmatrix}$$

Since $\lambda \to \lambda^*$, $\mu \to 0$. Thus Lagrangian methods lose the advantage of bounded parameters over barrier approaches.

Lancelot on our Example

Iteration	(New) Penalty Parameter	Trust Region Radius $\|\|\|\|_\infty$
16	1e-2	3.81 e-02
43	1e-4	1.1 e-02
85	1e-6	1.35 e-03
141	1e-8	4.22 e-05
203	1e-10	5.28 e-06
241	1e-12	1.70 e-06
268	1e-14	1.93
283	1e-16	4.41 e02
323	1e-18	2.19 e04
336	STOP	

Table 1: Reduction of the penalty parameter μ for LANCELOT

Observed Rate of Convergence for LINF

Iteration	$\frac{\phi(x^k)-\phi(x^*)}{\phi(x^{k+1})-\phi(x^*)}$
4	4.00
9	4.00
14	3.99
19	3.99
24	4.00
27	4.00

Table 2: Rates of convergence for the L_∞ penalty algorithm

Numerical Results

Nonlinear solver	$\|\|x^{final} - x^*\|\|_2$	Iterations	Message at termination
DONLP2	1.45e-16	4	Success
FilterSQP	5.26e-09	28	Convergence
LANCELOT	8.65e-07	336	Step size too small
LINF	1.05e-08	28	Step size too small
LOQO	1.60e-07	200	Iteration limit
LOQO	5.50e-07	1000	Iteration limit
MINOS	4.76e-06	27	Point cannot be improved
SNOPT	3.37e-07	3	Optimal Solution Found

Table 3: All tolerances set to 1e-16, except DONLP2

DONLP2 $<$ FSQP $<$ LINF $<$ LOQO $<$ SNOPT $<$ LANCELOT $<$ MINOS

Numerical runs observations

- Given the differences NLP solvers use as a measure for tolerance, the basis for comparison was the best achievable outcome (best shot).
- The fact that NLP solvers with augmented Lagrangian perform worse is somewhat expected, in light of our analysis.
- Note that LINF does only slightly worse than FSQP, though it does not use second-order information (nor it attempts to estimate it). This also shows that the problem is not in itself ill-conditioned.
- For FilterSQP, linear convergence was observed.
- For LOQO, increasing the number of iterations limit did not improve the results.
- Tolerances smaller than 10^{-16} may be a problem (LOQO). Some of the algorithms were well defined for 10^{-20} and the outcomes were almost identical with the ones for 10^{-16}.For tolerances in the range 10^{-12}–10^{-15} similar results are obtained.

SQP versus Interior-Point

- If a constraint is added twice, the minimizer (and the central path) of the original barrier $f(x) \Leftrightarrow \mu \ln(\Leftrightarrow g_1(x)) \Leftrightarrow \mu \ln(\Leftrightarrow g_2(x))$ shifts to satisfy

$$\nabla f(x_s(\mu)) \Leftrightarrow \frac{2\mu}{g_1(x_s(\mu))} \nabla g_1(x_s(\mu)) \Leftrightarrow \frac{\mu}{g_2(x_s(\mu))} \nabla g_2(x_s(\mu)) = 0$$

- However, the steepest descent QP has the same solution d even though the constraint is added twice:

$$\begin{array}{ll} \text{minimize} & \nabla f(x)^T d + \frac{1}{2} d^T d \\ \text{subject to} & g_j(x) + \nabla g_j(x)^T d \leq 0, \qquad j = 1, 2, 2 \end{array}$$

- Also, the penalty function $P(x) = \max\{g_0(x), g_1(x), ...g_m(x)\}$ is invariant to adding a constraint twice.

- Since the SQP is invariant to constraint repetition, it is reasonable to expect that it will be more robust than the interior point approach.

Conclusions

- We show that Quadratic Growth and MFCQ induce linear convergence of the L_∞ exact penalty method.
- We construct an example for which QG and MFCQ hold, but for which no locally convex augmented Lagrangian exists.
- We show that the SQP approach is more robust than Lagrangian methods, and possibly more robust than interior-point methods (for NLP).
- Any extension of these results would require unbounded multipliers, or some particularity of the constraint functions (convexity).
- The L_∞ penalty algorithm is not the answer when ill-conditioning is present (small maximum curvature on some of the critical cone directions). The problem of superlinear convergence under these assumptions is open.

Perturbation Theory

- Under the Traditional SOSC, a locally perturbed NLP will have a unique primal dual solution $(x(p), \lambda(p))$, which is Lipschitzian with respect to p.
- Under Robinson SOSC, the primal perturbed solution is unique $x(p)$, and Lipschitzian with respect to the perturbation. The dual solution is Lipschitzian at $p = 0$, $||\mathcal{M}(p) \Leftrightarrow \mathcal{M}(x^*)|| = O(||p||)$ (as sets).
- Under Shapiro SOSC, the primal perturbed solution is not necessarily unique and is Lipschitzian at x^* (as a set) with respect to the perturbation only for classes of perturbations (Maurer's example).