Continuous Optimisation Theorems

Lemma 1. *Let $f \in \mathcal{C}$, $x \in \mathbb{R}^n$, and $s \in \mathbb{R}^n$ with $s \neq 0$. Then*

$$\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x) \qquad \forall \alpha > 0 \text{ suff. small}$$

Lemma (Exact linesearch for quadratics)**.** *For $q(x) = g^T x + \frac{1}{2} x^T H x$, $\phi(\alpha) \coloneqq q(x + \alpha s)$,*

$$\alpha = -\frac{\nabla f(x)^T s}{s^T H s}$$

Definition (Armijo Condition)**.** *Choose $\beta \in (0, 1)$.*

$$f\left(x^k + \alpha^k s^k\right) \leq f\left(x^k\right) + \beta \alpha^k \nabla f\left(x^k\right)^T s^k$$

Lemma 2. *Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ with ∇f Lipschitz continuous with L, then Armijo condition at k^{th} satisfied for all $\alpha \in \left[0, \alpha^k_{\max}\right]$ where*

$$\alpha^k_{\max} = \frac{(\beta - 1)\nabla f\left(x^k\right)^T s^k}{L||s^k||^2}$$

Lemma 3. *Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ with ∇f Lipschitz continuous with L, then at k^{th} iteration,*

$$\alpha^k \geq \min\left\{\alpha_{(0)}, \tau \alpha^k_{\max}\right\}$$

Theorem 4 (Convergence of GLM)**.** *Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ bounded below by f_{low}, and ∇f Lipschitz continuous. Then either*

$$\exists l \geq 0 \text{ s.t. } \nabla f\left(x^l\right) = 0$$

or

$$\lim_{k \to \infty} \min \left\{ \frac{|\nabla f\left(x^k\right)^T s^k|}{||s^k||}, |\nabla f\left(x^k\right)^T s^k| \right\} = 0$$

Theorem 6 (Exact Linesearch Convergence)**.** *$f \in \mathcal{C}^2$, x^* local minimizer of f with $\nabla^2 f(x^*)$ positive definite between $\lambda^*_{\max}$ and $\lambda^*_{\min}$. With SD-e, if $x^k \to x^*$ as $k \to \infty$, then x^k converges linearly to x^*:*

$$\rho \leq \frac{\kappa(x^*) - 1}{\kappa(x^*) + 1} =: \rho_{SD}$$

where $\kappa(x^*) = \frac{\lambda^*_{\max}}{\lambda^*_{\min}} = \kappa\left(\nabla^2 f(x^*)\right)$.

Definition (Newton's Method)**.**

$$s^k \coloneqq -\left(\nabla^2 f\left(x^k\right)\right)^{-1} \nabla f\left(x^k\right)$$

Theorem 7 (Local Convergence of (Pure) Newton's Method)**.** • *$f \in \mathcal{C}^2(\mathbb{R}^n)$*

- *$\nabla f(x^*) = 0$*
- *$\nabla^2 f(x^*)$ nonsingular.*
- *$\nabla^2 f$ locally Lipschitz continuous at x^**

Then, $x^k \to x^*$ at quadratic rate.

Theorem 9 (Convergence of Newton bArmijo)**.** • *$f \in \mathcal{C}^2(\mathbb{R}^n)$ bounded below.*

- *∇f Lipschitz continuous.*
- *Newton's method + bArmijo linesearch.*
- *For all $k \geq 0$, eigenvalues of $\nabla^2 f(x^k)$ at iterates by positive and uniformly bounded below, away from zero, independently of k.*

Then either

$$\exists l \geq 0 \text{ s.t. } \nabla f\left(x^l\right) = 0$$

or

$$||\nabla f\left(x^k\right)|| \to 0$$

as $k \to \infty$

Definition (Secant Approximation)**.** *$B^k \approx \nabla^2 f\left(x^k\right)$ where*

$$\underbrace{\nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right)}_{\gamma^k} = B^{k+1} \underbrace{\left(x^{k+1} - x^k\right)}_{\delta^k}$$

Definition (SR1)**.** *$B^{k+1} := B^k + u^k \left(u^k\right)^T$ where*

$$u^k = \frac{\gamma^k - B^k \delta^k}{\rho^k}$$

where $\left(\rho^k\right)^2 := \left(\gamma^k - B^k\delta^k\right)^T \delta^k > 0$

Definition (BFGS)**.** *$B^{k+1} := B^k + u^k \left(u^k\right)^T + v^k \left(v^k\right)^T$ where*

$$u^k \left(u^k\right)^T = \frac{1}{\gamma^T \delta^k} \gamma^k \left(\gamma^k\right)^T$$

$$v^k \left(v^k\right)^T = -\frac{B^k \delta^k \left(B^k \delta^k\right)^T}{\left(\delta^k\right)^T B^k \delta^k}$$

where $\left(\rho^k\right)^2 := \left(\gamma^k - B^k\delta^k\right)^T \delta^k > 0$

Definition (Gauss-Newton)**.** *For nonlinear least-squares (NLS):*

$$f(x) := \frac{1}{2} \sum_{j=1}^{m} \left(r_j(x)\right)^2 = \frac{1}{2} ||r(x)||^2$$

where $r : \mathbb{R}^n \to \mathbb{R}^m$,

$$\nabla^2 f(x) = J(x)^T J(x) + \underbrace{\sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x)}_{Negligible} \textit{ suggests}$$

$$J\left(x^k\right)^T J\left(x^k\right) s^k = -J\left(x^k\right)^T r\left(x^k\right)$$

Definition (TR Decrease Param)**.** $\rho^k := \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - m_k(s^k)}$

Definition (Cauchy Point)**.** $\alpha_c^k := \arg\min_{\alpha > 0} m_k\left(-\alpha \nabla f\left(x^k\right)\right)$ *subject to* $||\alpha \nabla f\left(x^k\right)|| \leq \Delta_k$

Theorem 11 (GTR Global Convergence)**.**

- $f \in \mathcal{C}^2\left(\mathbb{R}^n\right)$ *bounded below.*
- ∇f *Lipschitz continuous.*
- $m_k\left(s^k\right) \leq m_k\left(s_c^k\right)$ *for all* k*.*

Then either

$$\exists k \geq 0 \textit{ s.t. } \nabla f\left(x^k\right) = 0$$

or

$$\lim_{k \to \infty} ||\nabla f\left(x^k\right)|| = 0$$

Theorem 12 (Cauchy Model Decrease)**.** *GTR with Cauchy decrease* $m_k\left(s^k\right) \leq m_k\left(s_c^k\right)$ *for all* k*,*

$$f\left(x^k\right) - m_k\left(s^k\right) \geq f\left(x^k\right) - m_k\left(s_c^k\right) \geq \frac{1}{2} ||\nabla f\left(x^k\right)|| \min\left\{\Delta_k, \frac{||\nabla f\left(x^k\right)||}{||\nabla^2 f\left(x^k\right)||}\right\}$$

Theorem 13 (Lower bound on TR radius)**.** $f \in \mathcal{C}^2\left(\mathbb{R}^n\right)$, ∇f *Lipschitz, Cauchy decrease. Suppose* $\exists \epsilon > 0$ *such that* $||\nabla f\left(x^k\right)|| \geq \epsilon$ *for all* k*, then*

$$\exists c \in (0, 1) \textit{ indep of } k \textit{ s.t. } \Delta_k \geq \frac{c}{L} \epsilon$$

Theorem 14 (At least one limit point is stationary)**.** $f \in \mathcal{C}^2\left(\mathbb{R}^n\right)$, ∇f *Lipschitz, Cauchy decrease. Then either* $\exists k \geq 0$ *s.t.* $\nabla f\left(x^k\right) = 0$ *or*

$$\liminf_{k \to \infty} ||\nabla f\left(x^k\right)|| = 0$$

Theorem 15 (Global Minimizer of TR Subproblem)**.**

$$\underbrace{(H + \lambda^* I)}_{positive\ semidef} s^* = -g$$

with $\lambda^* \geq 0$, $\lambda^*\left(||s^*|| - \Delta\right) = 0$, *and* $||s^*|| \leq \Delta$

Definition (KKT of (CP))**.**

$$\begin{aligned}
\nabla f\left(\hat{x}\right) &= J_E(x)^T \hat{y} + J_I(x)^T \hat{\lambda} \\
c_E\left(\hat{x}\right) &= 0 \\
c_I\left(\hat{x}\right) &\geq 0 \\
\hat{\lambda}_i &\geq 0 \\
\hat{\lambda}_i c_i\left(\hat{x}\right) &= 0
\end{aligned}$$

Definition (Lagrangian of (CP))**.**

$$\mathcal{L}(x, y, \lambda) := f(x) - y^T c_E(x) - \lambda^T c_I(x)$$
$$\nabla_x \mathcal{L}(x, y, \lambda) = \nabla f(x) - J_E(x)^T y - J_I(x)^T \lambda$$

so KKT implies $\nabla_x \mathcal{L}\left(\hat{x}, \hat{y}, \hat{\lambda}\right) = 0$

Theorem 16 (First Order Necessary Condition for (CP))**.** x^* *local minimizer implies* x^* *KKT under one of the conditions:*

- *Slater:* $\exists x$ *s.t.* $c_E(x) = Ax - b = 0$ *and* $c_I(x) > 0$
- *LICQ:* $\nabla c_i(x)$ *linearly indep.*

Definition (Convex Programming Problem)**.** $f(x)$ *is a convex function,* $c_i(x)$ *is a concave function for all* $i \in I$*, and* $c_E(x) := Ax - b$

Theorem 18 (Sufficient Optimality Conditions for Convex Problem)**.** *(CP) be convex programming problem, then KKT implies global minimizer.*

Theorem 19 (Second-order Necessary Conditions)**.** *(CP) satisfies some CQ. Let* x^* *be a local minimizer, and* (y^*, λ^*) *are Lagrange multipliers of KKT at* x^**.*

Then,

$$s^T \nabla^2_{xx} \mathcal{L}(x^*, y^*, \lambda^*) s \geq 0 \qquad \forall s \in F(\lambda^*)$$

Definition (Quadratic Penalty Function)**.**

$$\Phi_\sigma(x) = f(x) + \frac{1}{2\sigma} ||c(x)||^2$$

Theorem 21 (Global Convergence of Penalty Method)**.** *Apply basic quadratic penalty method. Assume* $f, c \in \mathcal{C}^1$*,* $y_i^k := -c_i\left(x^k\right)/\sigma^k$ *for* $i = 1, 2, \cdots, m$*, and*

$$||\nabla \Phi_{\sigma^k}\left(x^k\right)|| \leq \epsilon^k$$

where $\epsilon^k \to 0$*.*

Then, x^* *is KKT, and* $y^k \to y^*$*, the vector of Lagrange multipliers of constraints.*

Definition (Augmented Lagrangian)**.**

$$\Phi(x, u, \sigma) = f(x) - u^T c(x) + \frac{1}{2\sigma} ||c(x)||^2$$

Theorem 22 (Global Convergence of Augmented Lagrangian)**.** *Assuming* $f, c \in \mathcal{C}^1$ *in*

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } c(x) = 0$$

let $y^k = u^k - \frac{c(x^k)}{\sigma^k}$ *for some* $u^k \in \mathbb{R}^m$*, and assume that*

$$||\nabla \Phi\left(x^k, u^k, \sigma^k\right)|| \leq \epsilon^k$$

where $\epsilon^k \to 0, k \to \infty$*. Additionally assume* $x^k \to x^*$ *where* $\nabla c_i(x^*)$ *are linearly independent.*

Then, $y^k \to y^*$ *as* $k \to \infty$ *with* y^* *satisfying* $\nabla f(x^*) - J(x^*)^T y^* = 0$