

# 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

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

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

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

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

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

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

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

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

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

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

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

or

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

**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  $x^k \rightarrow x^*$  as  $k \rightarrow \infty$ , then  $x^k$  converges linearly to  $x^*$ :

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

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

$$s^k := -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

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

- $\nabla f$  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

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

or

$$\|\nabla f(x^k)\| \rightarrow 0$$

as  $k \rightarrow \infty$

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

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

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

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

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

**Definition** (BFGS). See Lecture Note.

**Definition** (Gauss-Newton).  $\nabla^2 f(x) = J(x)^T J(x) + \underbrace{\sum_{j=1}^m r_j(x) \nabla^2 r_j(x)}_{\text{Negligible}}$  suggests

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

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

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

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

Then either

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

or

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**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.

**Theorem 17** (Sufficient Optimality Conditions for Convex Problem). *(CP) be convex programming problem, then KKT implies global minimizer.*

See notes for second order conditions

**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  $f, c \in \mathcal{C}^1$ ,  $y_i^k := -c_i(x^k)/\sigma^k$  for  $i = 1, 2, \dots, m$ , and*

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

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

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