Constrained Optimality – Optimality Conditions

Lagrange's Theorem: $\boxed{L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x})}$

$\boxed{\nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* = 0}$

Lagrange Multipliers: $\frac{df}{dr}(\boldsymbol{x}(r)) = -\lambda_i$

Second Order Necessary Conditions: $\nabla_x \mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) = \nabla_x \boldsymbol{f}(\boldsymbol{x}^*) + \nabla_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* = 0$

$\boldsymbol{d}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d} \geq 0$ *for all* $\boldsymbol{d}$ *such that* $\nabla \boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{d} = 0$

Second Order Sufficient Conditions: $\boldsymbol{d}^T \nabla^2_{xx} \mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d} > 0$

KKT Theorem: $\boldsymbol{\mu}^* \geq 0$, $\nabla_x f(\boldsymbol{x}^*) + \nabla_x \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* + \nabla_x \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = 0$,

$\mu_i^* g_i(\boldsymbol{x}^*) = 0,\ i = 1,\ldots,p$, $h(\boldsymbol{x}^*) = 0,\ g(\boldsymbol{x}^*) \leq 0$

Second Order Necessary Conditions: $\boldsymbol{d}^\top \boldsymbol{H}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*,\boldsymbol{\mu}^*)\boldsymbol{d} \geq 0$

$$\boldsymbol{H}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = \nabla^2_{xx} \boldsymbol{f}(x) + \sum^m \lambda_i \nabla^2_{xx} \boldsymbol{h}_i(x) + \sum^p \mu_i \nabla^2_{xx} \boldsymbol{g}_i(x)$$

The Newton-Rhapson and Related Methods

Search direction: $\boldsymbol{d}_k = -\nabla^2 \boldsymbol{f}(\boldsymbol{x}_k)^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}_k) = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$

Exact line search

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \nabla^2 \boldsymbol{f}(\boldsymbol{x}_k)^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}_k)$$

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}_k - \alpha \nabla^2 \boldsymbol{f}(\boldsymbol{x}_k)^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}_k))$$

Convergence Theory: Positive Hessian

$m\boldsymbol{I} \preceq \nabla^2 \boldsymbol{f}(\boldsymbol{x})$, $\exists\ \eta > 0$ and $\theta > 0$

$\|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\|_2 > \eta \rightarrow f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}_k) \leq -\theta$

Levenberg–Marquardt Modification: If Hessian is not positive definite

$$\boldsymbol{d}_k = (\nabla^2 \boldsymbol{f}(\boldsymbol{x}_k) + \mu_k \boldsymbol{I})^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}_k)$$

Convexity

Line segment: $\{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{z} = \alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y},\ 0 \leq \alpha \leq 1\}$

Convex Functions: $f(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) \leq \alpha\ f(\boldsymbol{x}) + (1-\alpha)\ f(\boldsymbol{y}), \quad \forall \alpha \in (0,1)$

Strong Convexity: $MI \succeq \nabla^2 f(x) \succeq mI$

First Derivative Test: $f(\boldsymbol{x}) \geq f(\widehat{\boldsymbol{x}}) + \nabla f(\widehat{\boldsymbol{x}})^\top (\boldsymbol{x} - \widehat{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in C$

Second Derivative Test: $\boldsymbol{H}(\boldsymbol{x}) \succeq 0$ (positive semi-definite)

Mean Value Theorem: $f(b) = f(a) + \frac{df(x)}{dx}\Big|_{x=c}(b-a)$

$$f(\boldsymbol{b}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{x})^\top\Big|_{\boldsymbol{x}=\boldsymbol{a}}(\boldsymbol{b}-\boldsymbol{a}) + \frac{1}{2}(\boldsymbol{b}-\boldsymbol{a})^\top H(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{c}}(\boldsymbol{b}-\boldsymbol{a})$$

Constrained Optimality – Algorithms

Projected Gradient method: $\boldsymbol{x}_{k+1} = \Pi[\boldsymbol{x}_k - \alpha_k \boldsymbol{\nabla} f(\boldsymbol{x}_k)]$

Projection operator: $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{P}\boldsymbol{\nabla} f(\boldsymbol{x}_k)$$

Lagrangian algorithm:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k(\boldsymbol{\nabla} f(\boldsymbol{x}_k) + \boldsymbol{\nabla} \boldsymbol{h}(\boldsymbol{x}_k)\boldsymbol{\lambda}_k + \boldsymbol{\nabla} \boldsymbol{g}(\boldsymbol{x}_k)\boldsymbol{\mu}_k)$$
$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \beta_k \boldsymbol{h}(\boldsymbol{x}_k)$$
$$\boldsymbol{\mu}_{k+1} = P_+[\boldsymbol{\mu}_k + \gamma_k \boldsymbol{g}(\boldsymbol{x}_k)]$$

Penalty methods: $\min_x f(\boldsymbol{x}) + \gamma P(\boldsymbol{x})$. Penalty functions:

Abs. value	Courant-Beltrami	Logarithmic Barrier	Inverse Barrier
$\sum_{i=1}^{p} g_i^+(\boldsymbol{x})$	$\sum_{i=1}^{p} (g_i^+(\boldsymbol{x}))^2$	$-\sum_{i=1}^{p} \log(-g_i(\boldsymbol{x}))$	$-\sum_{i=1}^{p} \frac{1}{g_i(\boldsymbol{x})}$

First Order Methods

Descent direction: $\nabla f\left(\boldsymbol{x}^{(k)}\right)^\top \boldsymbol{d}^{(k)} < 0$

Steepest descent: $\boldsymbol{d}^{(k)} = -\nabla f\left(\boldsymbol{x}^{(k)}\right)$ $\alpha_k \in \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$

Backtracking line search: do $\alpha^{(k)} = c_\beta \alpha^{(k)}$ while

$$f\left(\boldsymbol{x}^{(k)} - \alpha^{(k)} \nabla f\left(\boldsymbol{x}^{(k)}\right)\right) > f(\boldsymbol{x}^{(k)}) - c_\alpha \alpha^{(k)} \left\|\nabla f\left(\boldsymbol{x}^{(k)}\right)\right\|_2^2$$

Convergence rate: $f\left(\boldsymbol{x}^{(k)}\right) - f^* \leq \left(1 - \frac{m}{M}\right)^k \left(f\left(\boldsymbol{x}^{(0)}\right) - f^*\right)$

Kantorovich Inequality: $\frac{(\boldsymbol{x}^T\boldsymbol{x})^2}{(\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x})(\boldsymbol{x}^T\boldsymbol{Q}^{-1}\boldsymbol{x})} \geq \frac{4\lambda_{\max}(\boldsymbol{Q})\lambda_{\min}(\boldsymbol{Q})}{(\lambda_{\max}(\boldsymbol{Q}) + \lambda_{\min}(\boldsymbol{Q}))^2} = \frac{4Mm}{(M+m)^2}$

Scaled Gradient Method: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{D}_k \nabla f\left(\boldsymbol{x}^{(k)}\right)$

One-dimensional Optimisation

Golden Section Search Method: $\varrho = \frac{3 \pm \sqrt{5}}{2}$

$$a_1 = a_0 + \varrho(b_0 - a_0), \; b_1 = a_0 + (1 - \varrho)(b_0 - a_0)$$

1-D Newton's Method:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Secant (Quasi-Newton) Method

$$x_{k+1} = x_k - f'(x_k)\frac{(x_k - x_{k-1})}{f'(x_k) - f'(x_{k-1})}$$

Lipschitz Continuity: $\|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq L \|\mathbf{x} - \mathbf{y}\|_2, \quad \forall \mathbf{x}, \mathbf{y} \in S$

Shubert's Algorithm: $\hat{x} = \frac{1}{2}\left(x_1 + x_2 + \frac{1}{L}\left(f(x_1) - f(x_2)\right)\right)$

Optimality Conditions

Feasible direction: $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$

First Order Necessary Condition: $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$

Second Order Necessary Condition: $\boldsymbol{d}^\top \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$

Second Order Sufficient Condition: $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ $\nabla^2 f(\boldsymbol{x}^*) \succ \boldsymbol{0}$

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top \left(\frac{\boldsymbol{A}^\top + \boldsymbol{A}}{2}\right)\boldsymbol{x}$$

Sylvester's criterion: all n upper left determinants positive