Optimization CHEATSHEET

Intro: formulation and basic analysis

1 Unconstrained optimization

1.1 Scalar (decision) variable

Minimize a scalar *cost function* over a real-valued variable. In mathematical lingo

$$\boxed{\underset{x\in\mathbb{R}}{\text{minimize}}\, f(x).}$$

The optimal cost is then

$$f^{\text{opt}} := \min_{x\in\mathbb{R}} f(x)$$

and the actual minimizing variable is

$$x^{\text{opt}} := \underset{x\in\mathbb{R}}{\text{argmin}}\, f(x)$$

Maximum trivially obtained as

$$\max f(x) = -\min(-f(x)).$$

1.1.1 First-order necessary

$$\boxed{\frac{\mathrm{d}f(x)}{\mathrm{d}x} = 0.}$$

1.1.2 Second order necessary

$$\boxed{\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \geq 0.}$$

1.1.3 Second-order sufficient condition of optimality

$$\boxed{\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} > 0.}$$

1.2 Vector (decision) variable

Minimize a scalar *cost function* over an n-tuple of real-valued variables. In mathematical lingo

$$\boxed{\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\, f(\mathbf{x}),}$$

where for computations $\mathbf{x}$ is viewed as a column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

1.2.1 First-order necessary condition of optimality

$$\boxed{\nabla f(\mathbf{x}) = \mathbf{0},}$$

where we view the gradient as a column vector

$$\nabla f(\mathbf{x}) := \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

Example: for $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + \mathbf{r}^{\mathrm{T}}\mathbf{x}$, the gradient is

$$\nabla f(\mathbf{x}) = \mathbf{Q}\mathbf{x} + \mathbf{r}$$

and the first-order condition of optimality is

$$\mathbf{Q}\mathbf{x} = -\mathbf{r}.$$

1.2.2 Second-order necessary condition of optimality

$$\boxed{\nabla^2 f(\mathbf{x}) \geq 0,}$$

where $\nabla^2 f(\mathbf{x})$ is the Hessian (the symmetrix matrix of the second-order mixed partial derivatives)

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix}$$

and the interpretation of the inequality ($\geq$) is not elememnt-wise but in terms of eigenvalues (the matrix is *positive semidefinite*).

Example: for $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + \mathbf{r}^{\mathrm{T}}\mathbf{x}$, the Hessian is

$$\nabla^2 f(\mathbf{x}) = \mathbf{Q}$$

and the first-order condition of optimality is

$$\mathbf{Q} \geq 0.$$

1.2.3 Second-order sufficient condition of optimality

$$\boxed{\nabla^2 f(\mathbf{x}) > 0,}$$

that is, the Hessian is *positive definite*.

2 Constrained optimization

2.1 Equality constraints

$$\boxed{\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} f(\mathbf{x}) \\ &\text{subject to } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \end{aligned}}$$

where $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^m$ defines a set of m equations

$$\begin{aligned} h_1(\mathbf{x}) &= 0 \\ h_2(\mathbf{x}) &= 0 \\ &\vdots \\ h_m(\mathbf{x}) &= 0. \end{aligned}$$

Lagrangian function (the original cost augmented with auxiliary Lagrange variable)

$$\boxed{\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) := f(\mathbf{x}) + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{h}(\mathbf{x}).}$$

2.1.1 First-order necessary condition of optimality

$$\boxed{\nabla \mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) = \mathbf{0},}$$

which amounts to two (generally vector) equations

$$\boxed{\begin{aligned} \nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\mathbf{x}) &= \mathbf{0} \\ \mathbf{h}(\mathbf{x}) &= \mathbf{0}. \end{aligned}}$$

Defining $\nabla \mathbf{h}(\mathbf{x}) \in \mathbb{R}^{n\times m}$ as stacked gradients (a tranpose of a Jacobian matrix)

$$\nabla \mathbf{h}(\mathbf{x}) := \begin{bmatrix} \nabla h_1(\mathbf{x}) & \nabla h_2(\mathbf{x}) & \dots & \nabla h_m(\mathbf{x}) \end{bmatrix},$$

the necessary condition can be written as

$$\boxed{\begin{aligned} \nabla f(\mathbf{x}) + \nabla \mathbf{h}(\mathbf{x})\boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{h}(\mathbf{x}) &= \mathbf{0}. \end{aligned}}$$

Beware of the nonregularity issue! The *Jacobian* $(\nabla \mathbf{h}(\mathbf{x}))^{\mathrm{T}}$ is regular at a given $\mathbf{x}$ (the $\mathbf{x}$ is a regular point) if it has a full column rank. Rank-deficiency reveals a defect in formulation.

Example:

$$\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + \mathbf{r}^{\mathrm{T}}\mathbf{x} \\ &\text{subject to } \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}. \end{aligned}$$

The first-order necessary condition of optimality is

$$\begin{bmatrix} \mathbf{Q} & \mathbf{A}^{\mathrm{T}} \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\mathbf{r} \\ \mathbf{b} \end{bmatrix}.$$

2.1.2 Second-order sufficient conditions

Using the unconstrained Hessian $\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda})$ is too conservative. Instead, use *projected Hessian*

$$\mathbf{Z}^{\mathrm{T}}\, \nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x},\boldsymbol{\lambda})\, \mathbf{Z} > 0,$$

where $\mathbf{Z}$ is an (orthonormal) basis of the nullspace of the Jacobian $(\nabla \mathbf{h}(\mathbf{x}))^{\mathrm{T}}$.

2.2 Inequality constraints—KKT conditions

$$\boxed{\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} f(\mathbf{x}) \\ &\text{subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \end{aligned}}$$

where $\mathbf{g}(\mathbf{x}) \in \mathbb{R}^p$ defines a set of p inequalities.

Karush-Kuhn-Tucker (KKT) conditions of optimality

$$\boxed{\begin{aligned} \nabla f(\mathbf{x}) + \sum_{i=1}^{p} \mu_i \nabla g_i(\mathbf{x}) &= \mathbf{0} \\ \mathbf{g}(\mathbf{x}) &\leq \mathbf{0} \\ \mu_i g_i(\mathbf{x}) &= 0, \quad i = 1,2,\dots,m \\ \mu_i &\geq 0, \quad i = 1,2,\dots,m. \end{aligned}}$$

2.2.1 Combination of equality and inequality constraints

$$\boxed{\begin{aligned} &\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} f(\mathbf{x}) \\ &\text{subject to } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ &\qquad\qquad \mathbf{g}(\mathbf{x}) \leq \mathbf{0}. \end{aligned}}$$

The KKT conditions

$$\boxed{\begin{aligned} \nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i \nabla g_i(\mathbf{x}) &= \mathbf{0} \\ \mathbf{h}(\mathbf{x}) &= \mathbf{0} \\ \mathbf{g}(\mathbf{x}) &\leq \mathbf{0} \\ \mu_i g_i(\mathbf{x}) &= 0, \quad i = 1,\dots,m \\ \mu_i &\geq 0, \quad i = 1,\dots,m. \end{aligned}}$$