Lecture 9, Oct 6, 2023

Optimization

Definition

An *optimization* problem in general seeks to minimize $f(\boldsymbol{x})$, subject to $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$ (equality constraints) and $\boldsymbol{h}(\boldsymbol{x}) \geq \boldsymbol{0}$ (inequality constraints), where $f : \mathbb{R}^n \mapsto \mathbb{R}, \boldsymbol{g} : \mathbb{R}^n \mapsto \mathbb{R}^m, \boldsymbol{h} : \mathbb{R}^n \mapsto \mathbb{R}^p$

- Note that $\operatorname{argmin} f(\boldsymbol{x}) = \operatorname{argmax}(-f(\boldsymbol{x}))$, so maximisation and minimization are interchangeable
- Some applications in robotics:
  - Motion planning: finding the most efficient path through a cluttered environment
  - Control: finding the sequence of inputs that cause a system to stay as close to a desired trajectory as possible
  - Machine learning: find a hypothesis/model that best explains the input data
  - Computer vision: matching an image to an existing map for localization
  - Reinforcement learning: optimizing a robot's performance over many training runs
- Example: steering a unicycle robot to follow the x axis by controlling ω_k with a constant v_k
  - Dynamics: $\underbrace{\begin{bmatrix} y_{k+1} \\ \theta_{k+1} \end{bmatrix}}_{\boldsymbol{x}_{k+1}} = \underbrace{\begin{bmatrix} y_k \\ \theta_k \end{bmatrix}}_{\boldsymbol{x}_k} + h \begin{bmatrix} v_k \sin\theta_k \\ \omega_k \end{bmatrix}$
  - Problem: given an initial condition (y_0, θ_0) find a sequence $\{\, \omega_0, \dots, \omega_K \,\}$ such that the cost function, $f = \sum_{k=0}^{K} (\boldsymbol{x}_k^T \boldsymbol{Q} \boldsymbol{x}_k + r\omega_k^2)$, is minimized
  - $\boldsymbol{Q}, r$ are the *optimization parameters*
    - In this problem, if we're following our desired trajectory, then the optimal $\boldsymbol{x}_k = \boldsymbol{x}^*$ for all k should be zero, therefore $\boldsymbol{x}_k$ is the error
    - $\boldsymbol{Q}$ is a matrix that weights each component of the error
    - The term $r\omega_k^2$ encourages the algorithm to make less aggressive turns
  - The decision variable is $\boldsymbol{\omega} = \begin{bmatrix} \omega_0 & \omega_1 & \cdots & \omega_K \end{bmatrix}^T$
  - The vehicle dynamics are the equality constraints
  - Inequality constraints can be e.g. $|\omega_k| \leq \omega_{\max}$
  - This is an example of *model predictive control*: finding the optimal set of control inputs to execute a trajectory for the next time period

Unconstrained Optimization

- *Unconstrained optimization* is optimization without equality or inequality constraints, i.e. minimizing only $\boldsymbol{f}(\boldsymbol{x})$

Definition

A point $\boldsymbol{x}^* \in \mathbb{R}^n$ is a *global minimum* of $f : \mathbb{R}^n \mapsto \mathbb{R}$ if

$$\forall \boldsymbol{x} \in \mathbb{R}^n, f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$$

$\boldsymbol{x}^*$ is a *local minimum* if

$$\exists \varepsilon > 0 \text{ s.t. } \forall \boldsymbol{x} \in \mathbb{R}^n, \|\boldsymbol{x} - \boldsymbol{x}^*\|_2 < \varepsilon \implies f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$$

- Assuming $f(\boldsymbol{x})$ is differentiable, then $\vec{\nabla} f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}^T \in \mathbb{R}^{1 \times n}$ describes the direction of steepest ascent; so $-\vec{\nabla} f$ is the direction of steepest descent

- If $\boldsymbol{x}_o$ is a local minimum, then $\vec{\nabla} f(\boldsymbol{x}_o) = \boldsymbol{0}$, which is a necessary but not sufficient condition for minimization

Definition

A *stationary point* of f is a point $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $\vec{\nabla} f(\boldsymbol{x}) = \boldsymbol{0}$.

Theorem

First-Order Optimality condition: If $\boldsymbol{x}$ is a local minimum of f, then $\boldsymbol{x}$ is a stationary point. Note that being a stationary point does not imply that $\boldsymbol{x}$ is a local minimum.

- A common strategy is to then find all stationary points $\boldsymbol{x}_i$, compute $f(\boldsymbol{x}_i)$ for all the points, and find the point with the lowest $f(\boldsymbol{x}_i)$
  - This is guaranteed to work, but it can be very hard to find all the local minima
  - Note we also need to check the boundaries with $\boldsymbol{x} \to \pm\infty$, e.g. if the function asymptotically converges

Definition

The *Hessian* matrix for a twice-differentiable f is the symmetric matrix

$$\boldsymbol{H}_f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

If (and only if) $\boldsymbol{x}^*$ is a stationary point, and $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is positive-definite (i.e. $\forall \boldsymbol{v}, \boldsymbol{v}^T \boldsymbol{H}_f \boldsymbol{v} > 0$, or that all eigenvalues are real and positive), then $\boldsymbol{x}^*$ is a local minimum. This is the *Second-Order Optimality Condition*.

- Note the Hessian reduces to a second derivative in the single-dimensional case
- This condition works because at a stationary point $f(\boldsymbol{x}) \approx f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \boldsymbol{H}_f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*)$ since the gradient disappears, so if $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is positive-definite, this expression is guaranteed to be greater than $f(\boldsymbol{x}^*)$
- Note also:
  - If $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is negative definite, then $\boldsymbol{x}^*$ is a local maximum
  - If $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is indefinite (positive and negative eigenvalues), then $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is a saddle point
  - If $\boldsymbol{H}_f(\boldsymbol{f}^*)$ is noninvertible (at least one zero eigenvalue), then odd things can happen, e.g. multiple local minima next to each other (flat function)

Definition

A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is *convex* if

$$\forall \boldsymbol{x}_1 \neq \boldsymbol{x}_2, \forall \alpha \in (0, 1), f((1 - \alpha)\boldsymbol{x}_1 + \alpha \boldsymbol{x}_2) \leq (1 - \alpha) f(\boldsymbol{x}_1) + \alpha f(\boldsymbol{x}_2)$$

f is said to be *strictly convex* if the $\leq$ is replaced with a $<$ in the above expression.

- Intuitively, this says that if we took two values $\boldsymbol{x}_1, \boldsymbol{x}_2$, then all the values of f in between the points will lie below the line connecting the two points

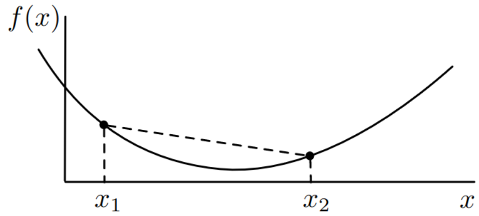

Figure 1: Illustration of convexity.

Theorem

A local minimum of a convex function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is necessarily a global maximum.

- This can be proven by contradiction
- f is convex if $\boldsymbol{H}_f(\boldsymbol{x})$ is positive definite for all $\boldsymbol{x}$
- Convex functions are much easier to work with, but in the real world, few functions are actually convex
  - Sometimes we can reformulate or relax specific parameters to make the function convex