descent algorithms

topic: (deterministic) gradient descent algorithms for unconstrained optimization

refs:

OPTIMAL CONTROL AND ESTIMATION

Robert F. Stengel

Department of Mechanical and Aerospace Engineering
Princeton University, Princeton, New Jersey

Section 3.6

Nonlinear Programming

SECOND EDITION

Dimitri P. Bertsekas

Massachusetts Institute of Technology

Chapter 1

- we've seen why derivatives can be useful to characterize (local) optima in

$$\text{(NLP)} \quad \min_{u \in \mathbb{R}^m} c(u)$$

← assume c is "smooth", i.e. continuously differentiable as many times as needed

- suppose we start with a "guess" $u \in \mathbb{R}^m$ and find $\underbrace{Dc(u)}_{\text{"gradient" of } c} \neq 0$

→ how should we modify our guess to get closer to a (local) min?

– regarding "$Dc(u)$" as a function $Dc(u) : \mathbb{R}^m \to \mathbb{R}$
$: v \mapsto \langle Dc(u), v \rangle$

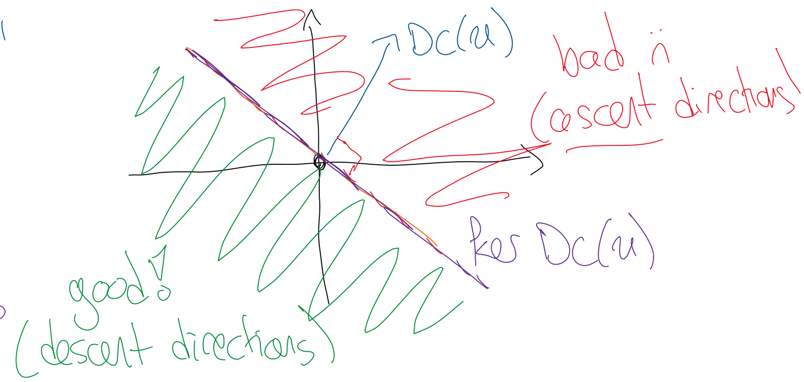
we see that the gradient defines a function that tells us, for any given direction $v \in \mathbb{R}^m$, how quickly c changes in the v direction:

$$c(u + \gamma \cdot v) \simeq c(u) + \gamma \cdot \langle Dc(u), v \rangle$$

* any $v \in \mathbb{R}^m$ s.t. $\langle Dc(u), v \rangle < 0$ is a valid descent direction

e.g. $v = -Dc(u) \Rightarrow \langle Dc(u), -Dc(u) \rangle = -\|Dc(u)\|^2 < 0$

but there's a whole "half-space" of descent directions!



among the infinite volume of possibilities

- to select a descent direction, we can formulate another optimization problem

→ solve $\min_{v \in \mathbb{R}^m} \langle Dc(u), v \rangle$ ← find the steepest / most rapid descent direction

s.t. $\|v\|_2 \leq \|Dc(u)\|_2$

– recalling that $\langle x, y \rangle = \|x\|_2 \|y\|_2 \cos\theta$,

θ = angle between x & y

$\Rightarrow$ solution is $v^* = -Dc(u)^T \in \mathbb{R}^m$

algorithm (gradient descent): $(*)\ u^+ = u - \gamma \cdot Dc(u)^T$, step size $\gamma > 0$

* note that $(*)$ defines a difference equation (DE)

→ observe/show that local minima of (NLP) are equilibria of (DE)

→ observe/ show that local minima of (NLP) are equilibria of (DE)

* let's use stability analysis of (DE) to find largest choice for $\gamma > 0$

→ linearize (DE) about a local min of (NLP)

- differentiating (*) wrt u: $D_u\left[u - \gamma \cdot Dc(u)^T\right] = \underbrace{I - \gamma \cdot D^2c(u)}_{A(\gamma)}$

  evaluating at u^* yields $(u-u^*)^+ \simeq A(\gamma)\cdot(u-u^*)$, $A(\gamma) =$

* need $|\lambda| < 1$ for all $\lambda \in \operatorname{spec} A(\gamma) = \operatorname{spec} I - \gamma \cdot D^2c(u)$

- spectral mapping theorem says that:

$$\text{if } \sigma \in D^2c(u) \text{ then } \lambda = 1 - \gamma\cdot\sigma \in \operatorname{spec} I - \gamma \cdot D^2c(u)$$

→ so we need $-1 < 1 - \gamma\cdot\sigma < +1 \iff 0 < \gamma < \frac{2}{\sigma}$, all $\sigma \in \operatorname{spec} D^2c(u)$

CHALLENGE

* → solve
$$\min_{v \in \mathbb{R}^m} \langle Dc(u), v\rangle$$
$$\text{s.t. } \|v\|_{D^2c(u)} \leq \|Dc(u)\|_{D^2c(u)}$$
$$\text{where } \|v\|_S := \sqrt{\tfrac{1}{2} v^T S v}$$

- let $g = Dc(u)$, $H = D^2c(u)$ so we're solving $\min_{v\in\mathbb{R}^m} \langle g, v\rangle$ s.t. $\|v\|_H \leq \|g\|_H$
$$\Updownarrow$$
$$\tfrac{1}{2} v^T H v \leq \tfrac{1}{2} g H g^T$$

- with $\tilde{c}(v,\lambda) = g\cdot v + \frac{\lambda}{2}\left(v^T H v - g H g^T\right)$

  then necessarily $D_v\tilde{c} = 0$, $D_\lambda \tilde{c} = 0$:

  - $D_v \tilde{c} = g + \lambda \cdot v^T H = 0 \iff v_0 = -\frac{1}{\lambda_0} H^{-1} g^T$

  - $D_\lambda \tilde{c} = \frac{1}{2}\left(v^T H v - g H g^T\right) = 0 \iff \|v_0\|_H^2 = \|g\|_H^2$

$$-D_x \tilde{C} = \frac{1}{2}\left(v^T H v - g H g^T\right) = 0 \iff \|v_0\|_H = \|g\|_H$$

$$\text{but } v_0 = -\frac{1}{\lambda_0} H^{-1} g^T \text{ so } \|v_0\|_H^2 = \frac{1}{\lambda_0^2} g H^{-1} H H^{-1} g^T = \frac{1}{\lambda_0^2} g H g^T = \frac{1}{\lambda_0^2} \|g\|_H^2$$

$$\text{so } \lambda_0^2 = 1 \iff |\lambda_0| = 1 \iff \lambda_0 = \pm 1$$

* sign of λ_0 determines whether v_0 is direction of <u>ascent</u> (−) or <u>descent</u> (+)

<u>note:</u> $u^+ = u - [D^2 c(u)]^{-1} Dc(u)^T$ is called Newton-Raphson