

Deciphering Jentzen's Lecture Notes

Jentzen's lecture notes, but *slightly* more human-readable with comment, and no fancy curly/Gothic letters you can't pronounce.

Chapter 4

This chapter contains mainly auxiliary results that should probably help to show convergence results of Gradient Descent in later chapters.

For clarity, we will assume differentiability and continuity of the derivative functions whenever necessary.

Lemma 4.2.1 (Gronwall's Inequality)

Let $T \in \mathbb{R}^+$ and $f : [0, T] \rightarrow \mathbb{R}$ satisfy

$$f'(t) \leq \alpha f(t) \tag{1}$$

for some $\alpha \in \mathbb{R}$. Then it holds for $t \in [0, T]$ that

$$f(t) \leq f(0)e^{\alpha t}$$

Proof. It suffices show that $r(t) := f(t)/e^{\alpha t} \leq f(0)$. Note that $r(0) = f(0)$, and that

$$r'(t) = \frac{f'(t) - \alpha f(t)}{e^{\alpha t}} \stackrel{(1)}{\leq} 0$$

The result follows from Fundamental Theorem of Calculus (FTC).

Lemma 4.2.3

This can be seen as a direct multivariate extension of Lemma 4.2.1.

Let $O \subset \mathbb{R}^d$ be an open set, and let $\theta_{t \in [0, T]}$ be a trajectory on O with some initial location θ_0 and velocity vector $g(\theta_t)$, i.e.,

$$\theta_t = \theta_0 + \int_0^t g(\theta_s) ds$$

If a function $V : O \rightarrow \mathbb{R}$ satisfies for all $\theta \in O$ that

$$\nabla V(\theta) \cdot g(\theta) \leq \alpha V(\theta) \tag{2}$$

then for all $t \in [0, T]$,

$$V(\theta_t) \leq V(\theta_0)e^{\alpha t}$$

Proof. Let $f(t) := V(\theta_t)$. By chain rule,

$$f'(t) = \nabla V(\theta_t) \cdot g(\theta_t) \stackrel{(2)}{\leq} \alpha V(\theta_t) = \alpha f(t)$$

Applying Gronwall's Inequality on f directly gives the result.

Corollary 4.2.5

Consider any cost function $J : \mathcal{O} \rightarrow \mathbb{R}$ we wish to minimize. In Gradient Descent, we take its negative gradient as the direction of each step $t \in \{0, \dots, N\}$, and get a discrete sequence of parameters $\theta_0, \theta_1, \dots, \theta_N$. Now, if we take really small step sizes, the sequence would be a pretty good approximation of a continuous trajectory, $\theta_{t \in [0, T]}$, where at any point in time the velocity of the trajectory is precisely the negative gradient of the cost function, $-\nabla J(\theta_t)$. (Jentzen also seems to have proved the uniqueness of such a trajectory. Check out 4.3.2 in notes if you're interested.)

With lemma 4.2.3, we can analyze how far the trajectory θ_t is from some θ^* under a **rather strong** assumption, that (2) holds with $V(\theta) = \|\theta - \theta^*\|_2^2$.

Let $\alpha > 0$. If for all $\theta \in \mathcal{O}$

$$\nabla J(\theta) \cdot (\theta - \theta^*) \geq \alpha \|\theta - \theta^*\|_2^2 \quad (3)$$

then it holds for all $t \in [0, T]$ that

$$\|\theta_t - \theta^*\|_2 \leq e^{-\alpha t} \|\theta_0 - \theta^*\|_2$$

Proof.

Define $g(\theta) := -\nabla J(\theta)$ and $V(\theta) := \|\theta - \theta^*\|_2^2$. We have

$$\nabla V(\theta) \cdot g(\theta) = -2(\theta - \theta^*) \cdot \nabla J(\theta) \stackrel{(3)}{\leq} -2\alpha \|\theta - \theta^*\|_2^2 = -2\alpha V(\theta)$$

Now apply lemma 4.2.3 and we get

$$\|\theta_t - \theta^*\|_2^2 = V(\theta_t) \leq V(\theta_0) e^{-2\alpha t} = e^{-2\alpha t} \|\theta_0 - \theta^*\|_2^2$$

Taking square roots on both sides gives

$$\|\theta_t - \theta^*\|_2 \leq e^{-\alpha t} \|\theta_0 - \theta^*\|_2$$

as desired.

Lemma 4.2.8, 4.2.9

(Proofs skipped)

How strong really is the assumption in lemma 4.2.5? Well, lemma 4.2.8 (we skip the proof) says if (3) holds in some closed ball B centered at θ^* , then for all $\theta \in B$,

$$J(\theta) - J(\theta^*) \geq \frac{\alpha}{2} \|\theta - \theta^*\|_2^2$$

Therefore θ^* is the **unique minimum** of J in B .

Hooray! Next let's make another strong assumption on J to get an upper bound of the difference $J(\theta) - J(\theta^*)$, because we can.

Let $\beta > 0$. Lemma 4.2.9 says if we assume

$$\|J(\theta)\|_2^2 \leq \beta \|\theta - \theta^*\| \quad (4)$$

for all θ in a closed ball B centered at θ^* , then

$$J(\theta) - J(\theta^*) \leq \frac{\beta}{2} \|\theta - \theta^*\|_2^2$$

for all $\theta \in B$.

Corollary 4.3.5

Pretty sure just a combination of the previous lemmas, with some technicalities involving $T \rightarrow \infty$

(Proof skipped)

Let $\alpha, \beta > 0, \theta^* \in \mathbb{R}^d, B$ be a closed ball centered at θ^* . If we pick initial location $\theta_0 \in B$, and let the trajectory θ_t be given by

$$\theta_t = \theta_0 - \int_0^t \nabla J(\theta_s) ds$$

Assume the followings hold for all $\theta \in B$:

$$\nabla J(\theta) \cdot (\theta - \theta^*) \geq \alpha \|\theta - \theta^*\|_2^2 \quad \|J(\theta)\|_2^2 \leq \beta \|\theta - \theta^*\|$$

Then we have for all $t \in [0, \infty)$

$$\|\theta_t - \theta^*\|_2 \leq e^{-\alpha t} \|\theta_0 - \theta^*\|_2$$

and

$$J(\theta) - J(\theta^*) \leq \frac{\beta}{2} e^{-2\alpha t} \|\theta_0 - \theta^*\|_2^2$$