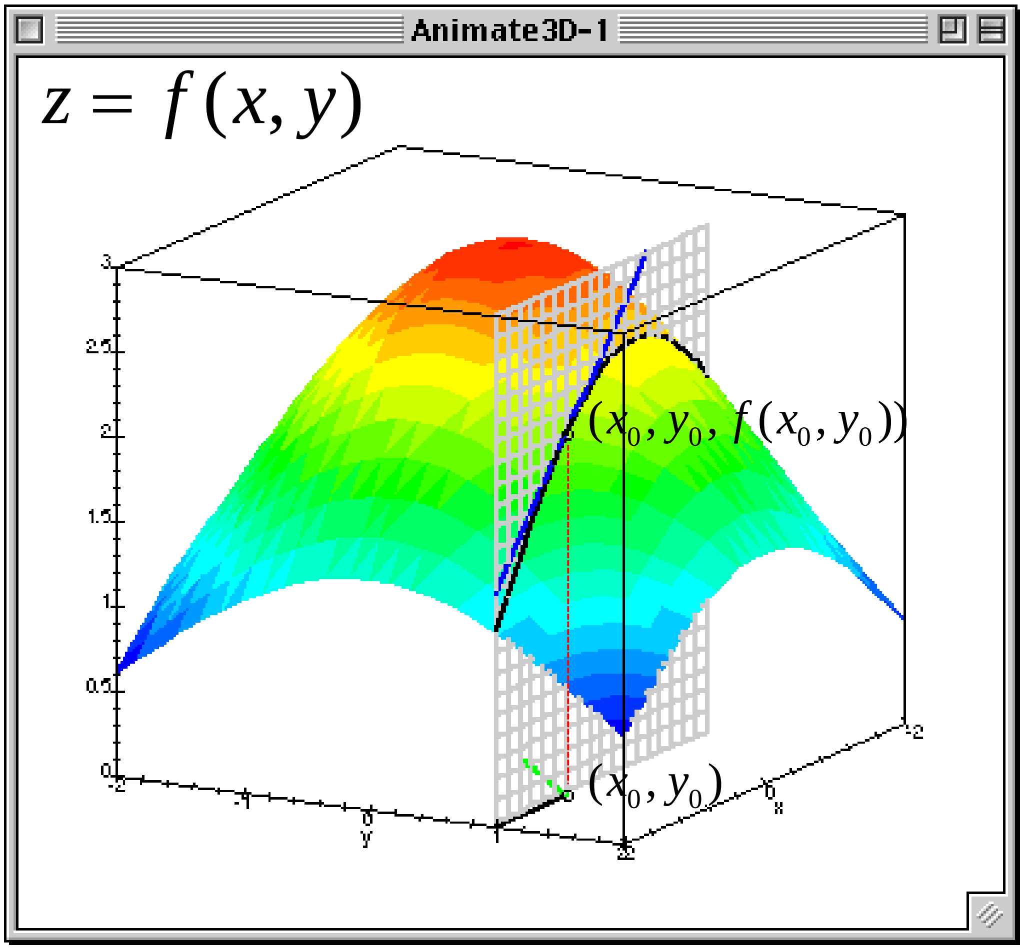
경사하강법 (Gradient Descent Method)

- ✓ 이변수함수의 방향도함수
- ✓ 이변수함수의 그래디언트
- ✓ 경사하강법의 개념

방향도함수의 의미 $D_{\vec{u}=(u_1,u_2)} f(x_0, y_0)$



: $\vec{u} = (u_1, u_2)$

방향으로의 파란색 직선의 기울기 :

$$D_{\vec{u}=(u_1,u_2)} f(x_0, y_0)$$

$$= \lim_{h \to 0} \frac{f((x_0, y_0) + h(u_1, u_2)) - f((x_0, y_0))}{h}$$

방향도함수의 성질과 그래디언트 벡터

- $z = f(x, y)$가 (x_0, y_0)에서 미분가능할 때 (x_0, y_0)에서 $\vec{u} = (u_1, u_2)$방향으로 방향도함수는

$$
\begin{aligned}
& D_{\vec{u}=(u_1,u_2)} f(x_0, y_0) \\
& = f_x(x_0, y_0)u_1 + f_y(x_0, y_0)u_2 \\
& = \nabla f(x_0, y_0) \cdot \vec{u}
\end{aligned}
$$

- $z = f(x, y)$가 (x_0, y_0)에서 gradient vector

$$\nabla f(x_0, y_0) = (f_x(x_0, y_0), f_y(x_0, y_0))$$

그래디언트의 성질

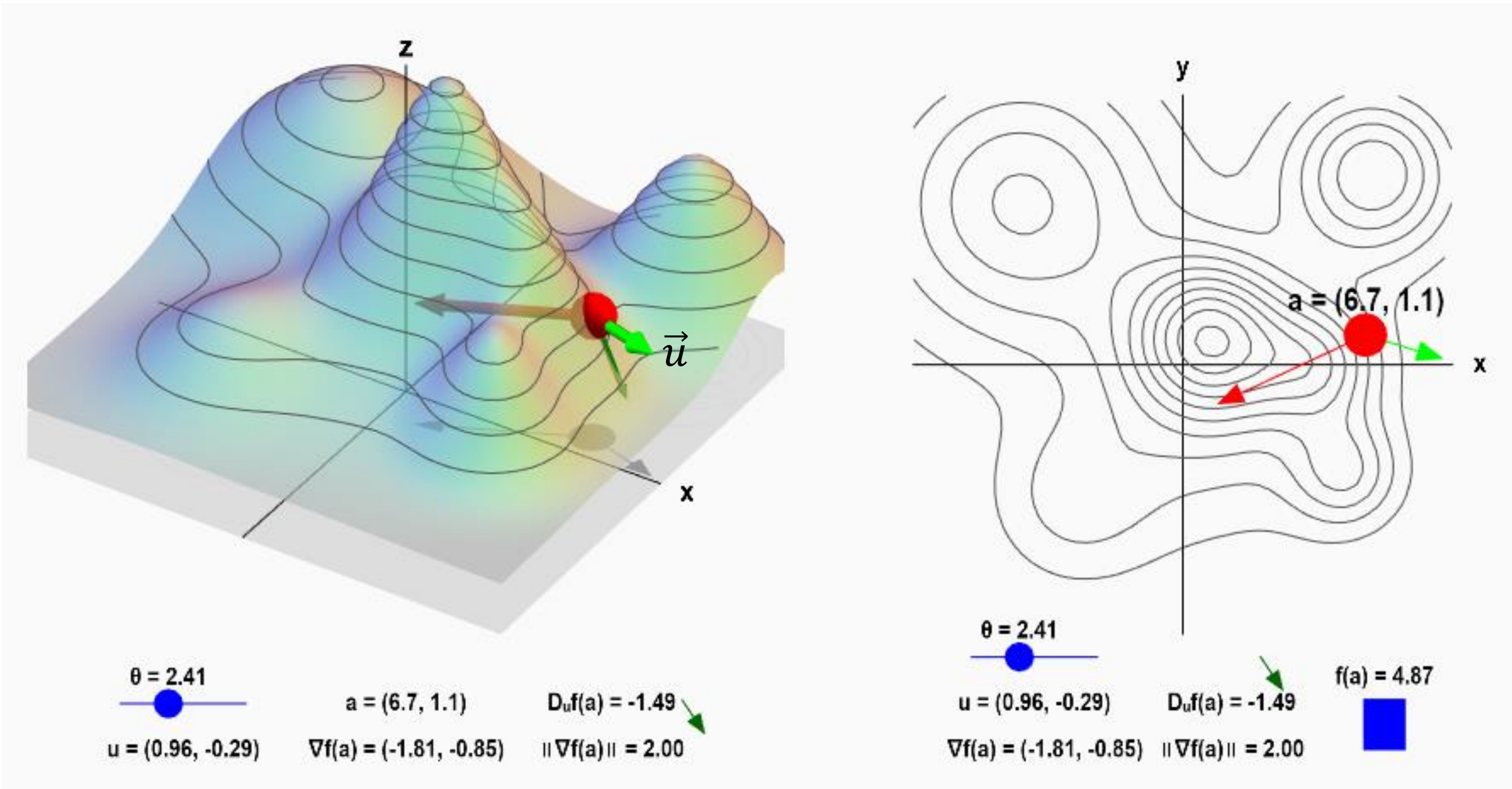
정리 11.11 그래디언트의 성질

함수 f가 점 (x, y)에서 미분가능이라고 하면

1. $\nabla f(x, y) = \mathbf{0}$이면 모든 $\mathbf{u}$에 대하여 $D_{\mathbf{u}} f(x, y) = 0$이다.
2. f의 최대증가의 방향은 $\nabla f(x, y)$로 주어지고, $D_{\mathbf{u}} f(x, y)$의 최댓값은 $\|\nabla f(x, y)\|$이다.
3. f의 최소증가의 방향은 $-\nabla f(x, y)$로 주어지고, $D_{\mathbf{u}} f(x, y)$의 최솟값은 $-\|\nabla f(x, y)\|$이다.

- 함수 $f(x, y)$가 어느 방향으로 가장 빠르게 증가하는가?
  - 가장 빠른 상승 방향
  - 그래디언트로 주어짐

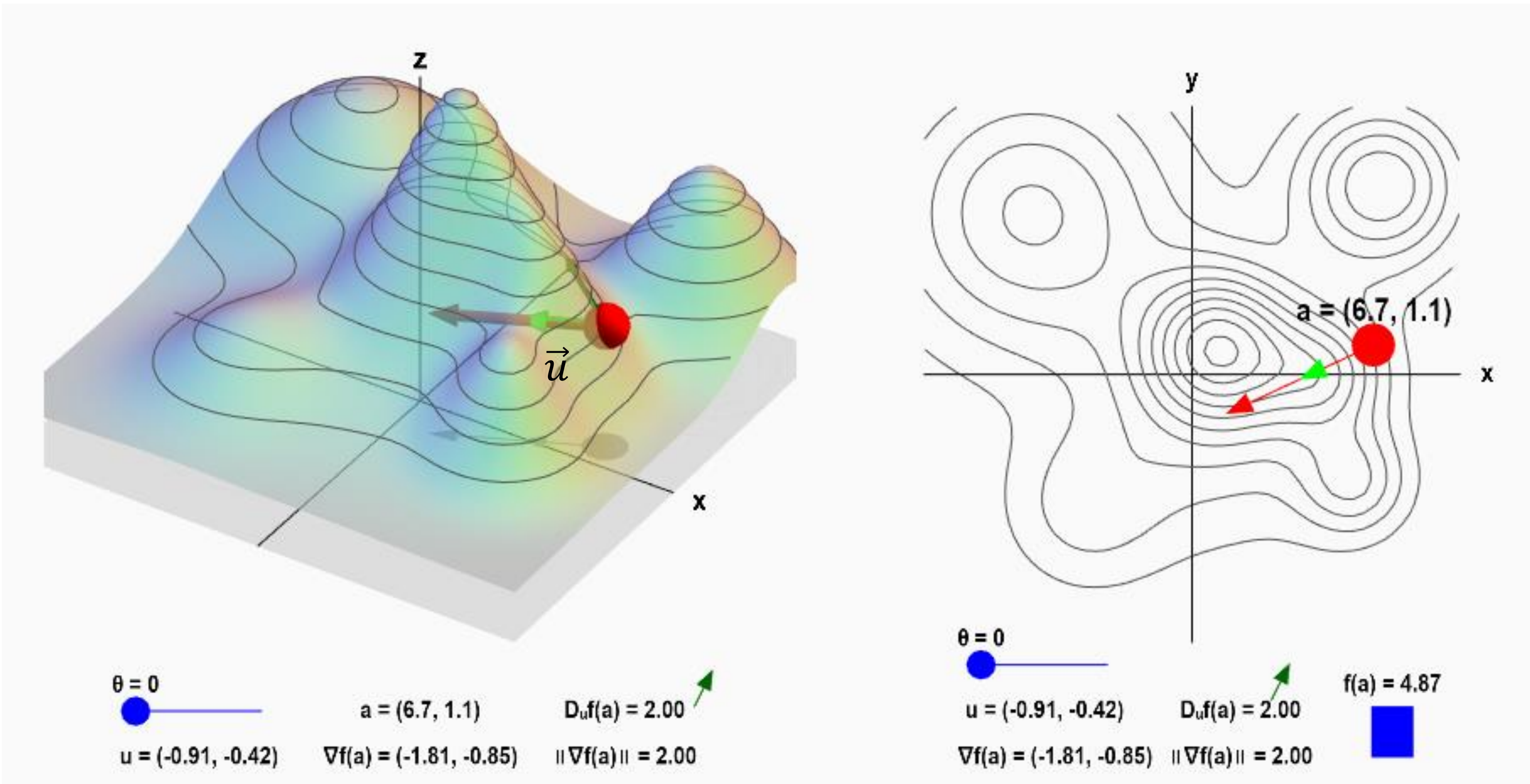
그래디언트의 성질



z
x
$\vec{u}$
θ = 2.41
u = (0.96, -0.29)
a = (6.7, 1.1)
∇f(a) = (-1.81, -0.85)
D_uf(a) = -1.49
‖∇f(a)‖ = 2.00
y
x
a = (6.7, 1.1)
θ = 2.41
u = (0.96, -0.29)
∇f(a) = (-1.81, -0.85)
D_uf(a) = -1.49
‖∇f(a)‖ = 2.00
f(a) = 4.87

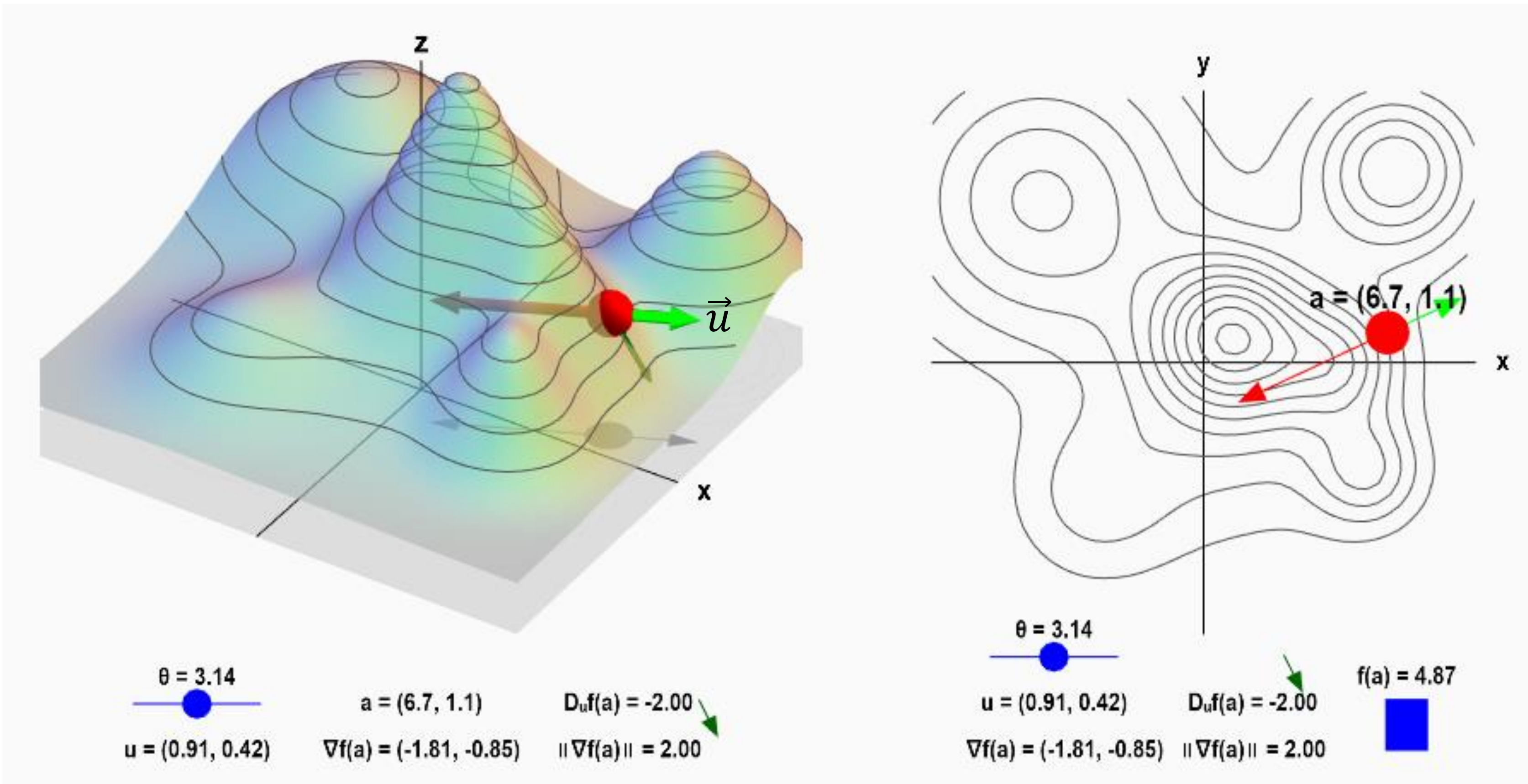
그래디언트의 성질

$\vec{u}$가 grad f(a) 방향 방향벡터일 때



그래디언트의 성질

$\vec{u}$가 grad f(a) 의 정반대 방향 방향벡터일 때



경사하강법의 개념

The simplest algorithm in the world (almost). Goal:

$$\underset{x}{\text{minimize}} \;\; f(x)$$

Just iterate

$$x_{t+1} = x_t - \eta_t \nabla f(x_t)$$

where η_t is stepsize.

- 최대값 또는 극대값을 구하고 싶으면

  $x_{t+1} = x_t + \eta_t \nabla f(x_t)$

- η_t는 learning rate라고도 부르며 t에 따라 변화시키지 않고 상수로 쓰는 경우도 많이 있음