Convex Programming
Problem Set 5 – CS 6515/4540 (Fall 2025)

Solutions to problem statement 5, Q19.

19 Gradient Descent Failure

For $f(x) = \frac{1}{2}x^2$ the gradient is $f'(x) = x$. Gradient descent with step size $\eta > 0$ updates

$$x_{t+1} = x_t - \eta f'(x_t) = (1 - \eta)x_t.$$

Thus, for any initial x_0,

$$x_t = (1 - \eta)^t x_0.$$

Set $r := 1 - \eta$. If $|r| > 1$ then $|x_k| \to \infty$. This occurs exactly when $\eta > 2$ for positive η.
For example $\eta = 3$ and $x_0 = 1$. Then $r = 1 - \eta = -2$ and

$$x_k = (-2)^k.$$

The t-th average is

$$\overline{x}_t := \frac{1}{t}\sum_{k=0}^{t-1} x_k = \frac{1}{t}\sum_{k=0}^{t-1}(-2)^k = \frac{1 - (-2)^t}{3t},$$

so $|\overline{x}_t| \sim \dfrac{|(-2)^t|}{3t} \to \infty$ as $t \to \infty$. Therefore the time-averages do not converge to the optimum 0 (in fact they diverge) as $t \to \infty$.
More generally, for any $\eta > 2$ and any $x_0 \neq 0$ we have $|1 - \eta| > 1$ and the time-averages $\overline{x}_t$ diverge, so the averaged iterates cannot converge to the minimizer.