

A Short Note on Gradient Descent for Strongly Convex Quadratics

A. Example*

October 2026

Abstract

We restate the classical linear convergence result for gradient descent on strongly convex quadratic functions, give the optimal constant step size, and show how the number of iterations grows with the condition number. The aim is a compact, self-contained reference with the proof, the algorithm and the resulting iteration counts.

1 Setting

Let $f(x) = \frac{1}{2}x^\top Ax - b^\top x$ with $A \in \mathbb{R}^{n \times n}$ symmetric and

$$\mu I \preceq A \preceq LI, \quad 0 < \mu \leq L. \quad (1)$$

Definition 1. The *condition number* of f is $\kappa = L/\mu \geq 1$.

The unique minimiser is $x^\star = A^{-1}b$. Gradient descent with constant step $\eta > 0$ is Algorithm 1.

Algorithm 1 Gradient descent with constant step

Require: $x_0 \in \mathbb{R}^n$, step $\eta > 0$, tolerance $\varepsilon > 0$

```
1:  $k \leftarrow 0$ 
2: while  $\|\nabla f(x_k)\| > \varepsilon$  do
3:    $x_{k+1} \leftarrow x_k - \eta (Ax_k - b)$ 
4:    $k \leftarrow k + 1$ 
5: end while
6: return  $x_k$ 
```

2 Convergence

Theorem 1. Under (1), the step $\eta = 2/(L + \mu)$ gives

$$\|x_k - x^\star\| \leq \left(\frac{\kappa - 1}{\kappa + 1} \right)^k \|x_0 - x^\star\|. \quad (2)$$

Proof. Since $\nabla f(x^\star) = 0$, the error $e_k = x_k - x^\star$ satisfies $e_{k+1} = (I - \eta A) e_k$. The eigenvalues of $I - \eta A$ are $1 - \eta\lambda$ with $\lambda \in [\mu, L]$, so $\|e_{k+1}\| \leq \max_{\lambda \in [\mu, L]} |1 - \eta\lambda| \|e_k\|$. The maximum is minimised when $1 - \eta\mu = -(1 - \eta L)$, that is $\eta = 2/(L + \mu)$, where it equals $(L - \mu)/(L + \mu) = (\kappa - 1)/(\kappa + 1)$. Induction on k gives (2). $\square$

*Illustrative example generated with latex-ops.

3 Iteration counts

By (2), reducing the error by a factor 10^{-6} takes the smallest k with $\left((\kappa - 1)/(\kappa + 1)\right)^k \leq 10^{-6}$. Figure 1 plots the bound and Table 1 lists the counts; for large κ they approach $\frac{\kappa}{2} \ln 10^6$, linear in the condition number.

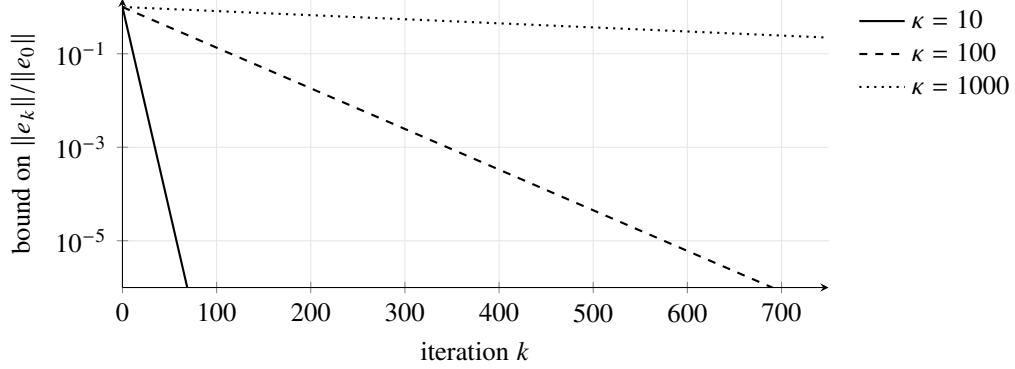


Figure 1: The bound (2) for three condition numbers. Line style, not colour, distinguishes the curves, so the figure survives grayscale printing.

Table 1: Iterations needed to reduce the error by 10^{-6} with $\eta = 2/(L + \mu)$.

κ	rate $(\kappa - 1)/(\kappa + 1)$	iterations	$\lceil \frac{\kappa}{2} \ln 10^6 \rceil$
10	0.8182	69	70
100	0.9802	691	691
1000	0.9980	6908	6908

4 Remarks

The bound is tight for quadratics: an initial error along the eigenvectors of μ or L attains it. Methods with momentum improve the dependence from κ to $\sqrt{\kappa}$; that comparison is outside the scope of this note.