

Optimization

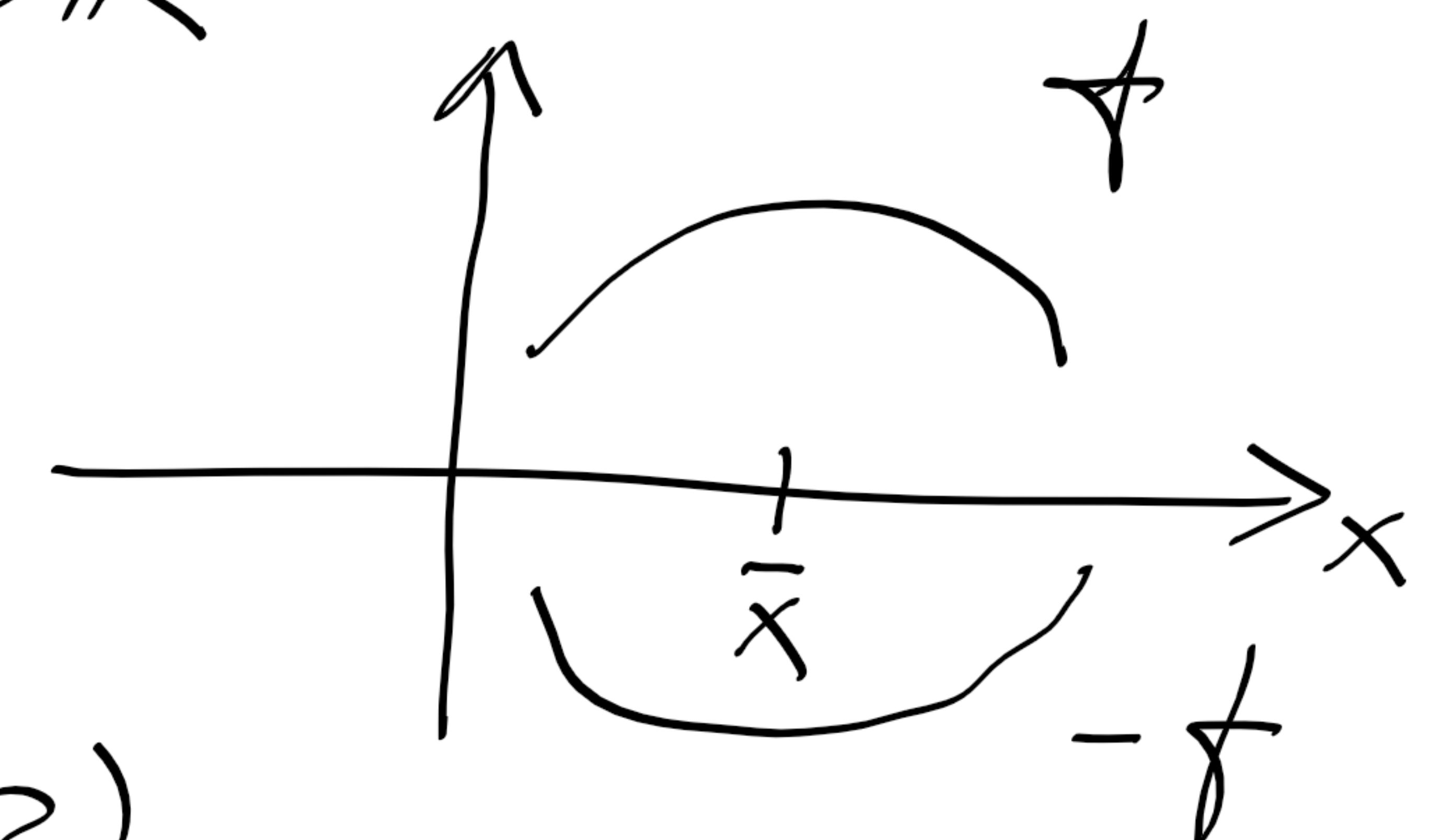
Problem

$$(P) \quad \text{minimize } f(x)$$

subject to $x \in S \subseteq \mathbb{R}^n$

where the objective function $f: S \rightarrow \mathbb{R}$

Note: $\max f = -\min(-f)$



* Unconstrained optimization (Ch. 2-3)

$S = \mathbb{R}^n$ algorithms to find an approximate minimizer

* Constrained optimization (Ch. 4-8)

$$S = \left\{ x \in \overline{X} \subseteq \mathbb{R}^n : \begin{array}{l} g_i(x) \leq 0, \quad i=1, \dots, m, \\ h_j(x) = 0, \quad j=1, \dots, l \end{array} \right\}$$

KKT theory, convexity, duality

* Penalty/barrier functions (Ch. 9):

Convert constrained non-linear (P) to unconstrained and use algorithms.

Notation

- Def. $x \in \mathbb{R}^n$ is **feasible** iff $x \in S$.
A constraint $g_i(x) \leq 0$ is **active** at $\bar{x} \in \mathbb{R}^n$ iff $g_i(\bar{x}) = 0$.
- vector $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$.
- scalar product $x^T y = (x_1 \dots x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$
 $= \sum_{i=1}^n x_i y_i = y^T x$

- length $\|x\| = \sqrt{x^T x} = \sqrt{\sum x_i^2}$

- gradient $\nabla f = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$

- Hessian

$$\nabla^2 f = \nabla \nabla^T f = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \left(\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} \right) f$$

$$= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Symmetric
when $f \in C^2$

- we write $\nabla f^T = (\nabla f)^T = \nabla^T f$
- A line in \mathbb{R}^n through $x_0 \in \mathbb{R}^n$ with direction d : $x = x_0 + t d$
↑
parameter

Review of calculus

- Behaviour of f along the line $x(t) = x_0 + t d$ is given by

$$F(t) = f(x(t)) = f(x_0 + t d)$$

$$F'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} = \sum_i \frac{\partial f}{\partial x_i} d_i = \nabla f^T d = d^T \nabla f$$

$$F''(t) = \frac{d}{dt} F'(t) = \underline{\underline{d^T \nabla^2 f^T d}} = d^T \nabla \nabla^T f d = d^T \nabla^2 f d$$

Hessian \uparrow

• Directional derivative: $\|\mathbf{d}\| = 1$

$$f'(x_0; \mathbf{d}) = \lim_{t \rightarrow 0} \frac{f(x_0 + t\mathbf{d}) - f(x_0)}{t} = \lim_{t \rightarrow 0} \frac{F(t) - F(0)}{t}$$

$$= F'(0) = \nabla f(x_0)^T \mathbf{d}$$

Property: $f'(x_0; \mathbf{d}) = \nabla f(x_0)^T \mathbf{d} \leq \|\nabla f(x_0)^T \mathbf{d}\|$
 $\leq \|\nabla f(x_0)\| \|\mathbf{d}\| = \|\nabla f(x_0)\|$

↑ Cauchy-Schwarz
inequality

with equality iff $\nabla f(x_0) = t\mathbf{d}$,
for some $t \geq 0$.

• Proposition: $\nabla f(a) \perp$ level surface

$$f(x) = C$$

through $a \in \mathbb{R}^n$

Proof: True if $\nabla f(a) = 0$. Otherwise we

$\frac{\partial f(a)}{\partial x_n} \neq 0$. Implicit function then gives

$$g \in C^1 : f(x) = C \Leftrightarrow$$

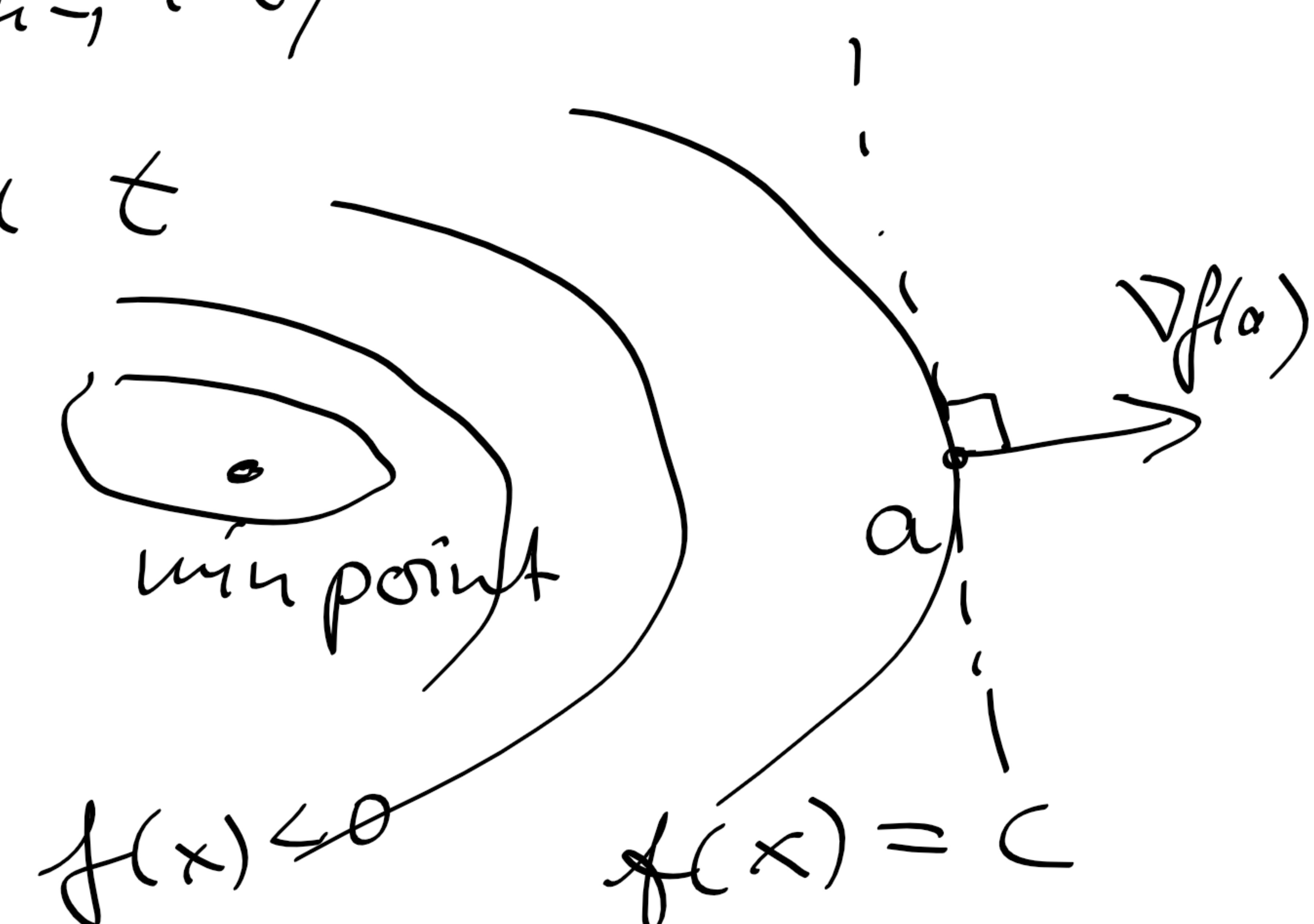
$$x_n = g(x_1, x_2, \dots, x_{n-1}) \quad \text{near } a$$

$$\Leftrightarrow \begin{cases} x_1 = a_1 + t \\ x_2 = a_2 + t \\ \vdots \\ x_{n-1} = a_{n-1} + t \\ x_n = g(a_1 + t, \dots, a_{n-1} + t) \end{cases} \quad \text{small } t$$

$$\Leftrightarrow x = \hat{x}(t) \quad \text{small } t$$

$$\text{Then } f(\hat{x}(t)) = C \Rightarrow$$

$$\nabla f(a)^T \underbrace{\hat{x}'(0)}_{\text{tangent}} = 0$$



- Taylor expansion of $f \in C^2$ at $x_0 \in \mathbb{R}^n$:

with $F(t) = f(x_0 + t\alpha)$

$$F(t) = F(0) + t F'(0) + \frac{1}{2} t^2 F''(0) + o(t^2)$$

where $\frac{o(t)}{t} \rightarrow 0$ as $t \rightarrow 0$

$$f(x_0 + t\alpha) = f(x_0) + t \alpha^T \nabla f(x_0) + \frac{1}{2} t^2 \alpha^T \nabla^2 f(x_0) \alpha + o(t^2)$$

$$\text{Set } x = x_0 + t\alpha \Leftrightarrow t\alpha = x - x_0 + o(t^2)$$

$$f(x) = f(x_0) + (x - x_0)^T \nabla f(x_0) + \frac{1}{2} (x - x_0)^T \nabla^2 f(x_0) (x - x_0) + o(\|x - x_0\|^2)$$

Conditions for local minima

$$(P) \quad \underset{x \in S}{\text{minimize}} \quad f(x)$$

Def. $\bar{x} \in S$ is a

- global minimizer if $f(x) \geq f(\bar{x}) \quad \forall x \in S$
- local minimizer if ————— and $\|x - \bar{x}\| < \delta$ for some δ
- strict local min. if $f(x) > f(\bar{x})$ when $x \neq \bar{x}$

Let $S = \mathbb{R}^n$

Thm (necessary conditions)

\bar{x} local minimizer of $f \in C^2 \Rightarrow$

$$\begin{cases} \nabla f(\bar{x}) = 0 \\ \nabla^2 f(\bar{x}) \text{ positive semidef.} \end{cases}$$

Proof: Let $F(t) = f(\bar{x} + t\alpha)$, α arbitrary

Slope $0 = F'(0) = \alpha^T \nabla f(\bar{x}) \quad \forall \alpha \Rightarrow \nabla f(\bar{x}) = 0$

Curvature $0 \leq F''(0) = \alpha^T \nabla^2 f(\bar{x}) \alpha \quad \forall \alpha \quad \#$

Thm (sufficient conditions)

$f \in C^2$, $\nabla f(\bar{x}) = 0$, $\nabla^2 f(\bar{x})$ pos. def.

$\Rightarrow \bar{x}$ strict local minimizer.

Proof: Taylor expansion with $\alpha \neq 0$:

$$\begin{aligned} f(\bar{x} + t\alpha) - f(\bar{x}) &= \frac{1}{2} t^2 \alpha^T \nabla^2 f(\bar{x}) \alpha + o(t^2) \\ &= t^2 \left(\underbrace{\frac{1}{2} \alpha^T \nabla^2 f(\bar{x}) \alpha}_{\text{fixed scalar} > 0} + o(1) \right) > 0 \quad \text{for small } t \quad \# \end{aligned}$$

$\underbrace{\alpha^T \nabla^2 f(\bar{x}) \alpha}_{\text{fixed scalar} > 0} \rightarrow 0, t \rightarrow 0$

Very important to study optimization algorithms
on a quadratic function:

$$q(x) = \frac{1}{2} x^T H x + c^T x + b$$

$$\nabla q(x) = Hx + c \quad (\text{Exercise!})$$

$$\nabla^2 q(x) = H$$

Unconstrained optimization

minimize $f(x)$
 $x \in \mathbb{R}^n$

Prototype algorithm: Given a starting point x_0 , do for $k=1, \dots$

- given x_k , choose a search direction d_k
- find $\lambda_k \in \mathbb{R}$ that minimizes

$$F(\lambda) = f(x_k + \lambda d_k) \quad (\text{line search})$$

- $x_{k+1} = x_k + \lambda_k d_k$
- stop criterion (see p. 108)

Def. d_k is a **descent direction** at x_k
iff $f(x_k + \lambda d_k) < f(x_k)$ when $0 < \lambda < \delta$
for some δ

Subproblems

- choice of d_k — choice of method
- choice of line search and how accurate
- guarantee for convergence (Ch. 10)
generally only to a stationary point
- convergence rate (App. 1, Ch. 3)

Line search (Chapter 2)

Problem: minimize $F(\lambda)$
 $a \leq \lambda \leq b$

Assume F is *unimodal* with minimizer $\bar{\lambda}$.

Algorithms:

- without derivatives

Dichotomous search

Golden section search

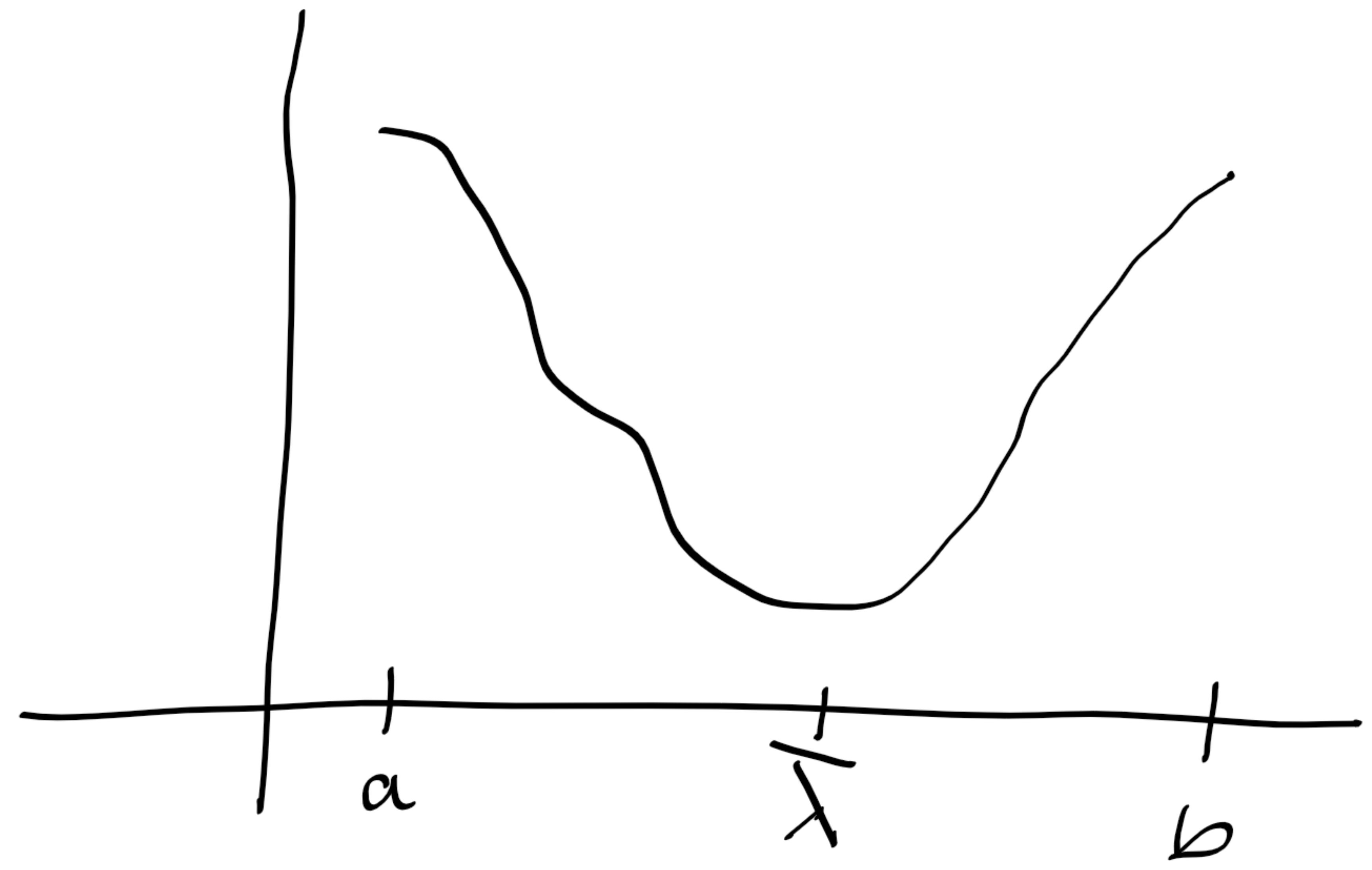
Quadratic fit $F(\lambda_k), F(\lambda_{k+1}), F(\lambda_{k+2})$

- with derivatives

Bisection

Armijo's rule

Quadratic fit using $F(\lambda_k), F(\lambda_{k+1}), F'(\lambda_k)$



Dichotomous search

Initial interval length

$$L_0 = b - a.$$

New interval after two function evaluations.

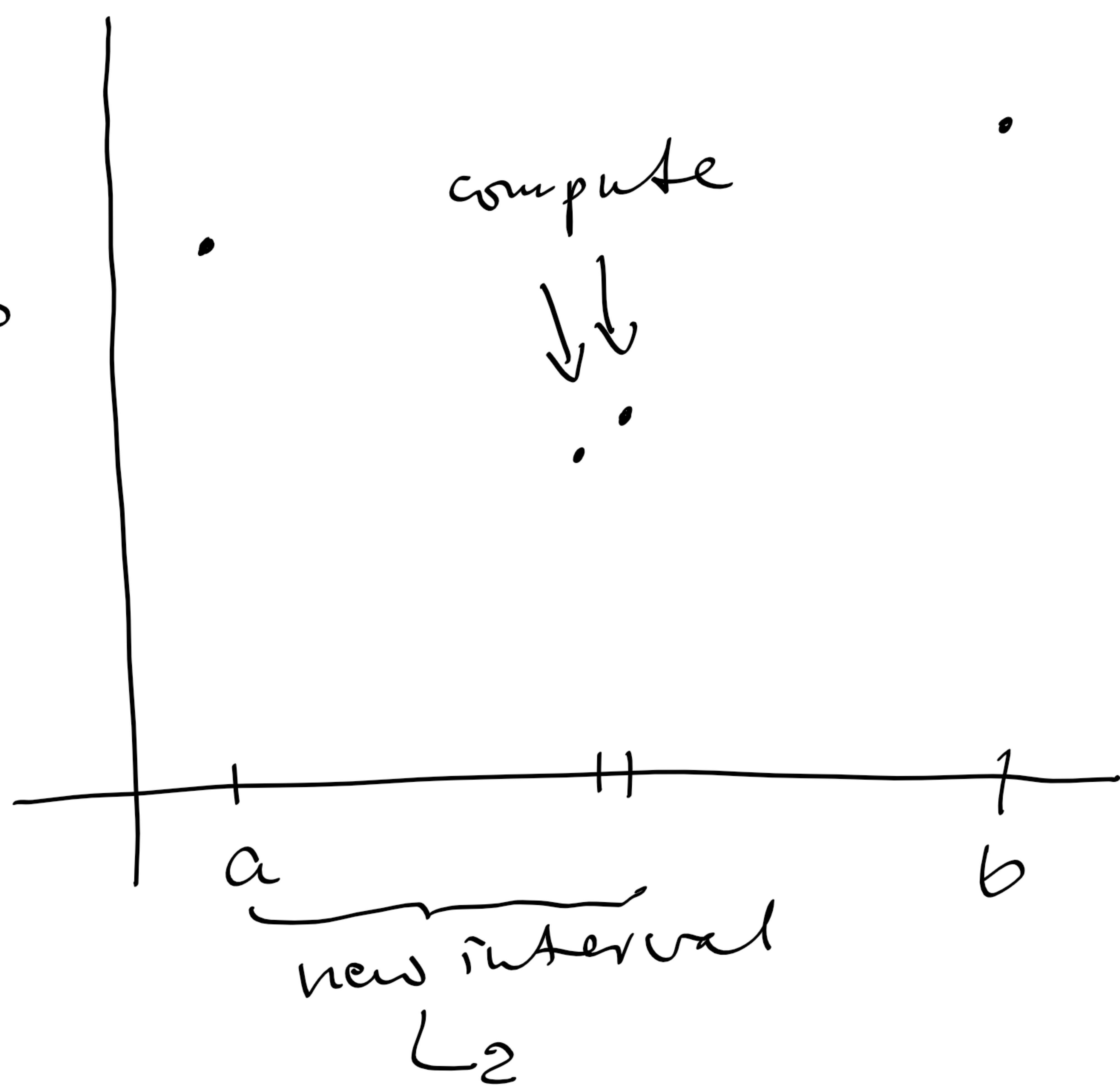
$$L_2 \approx \frac{1}{2} L_0$$

$$L_4 \approx \frac{1}{2} L_2 = \frac{1}{2^2} L_0$$

$$L_N \approx \frac{1}{2^{N/2}} L_0 = \left(\frac{1}{\sqrt{2}}\right)^N L_0$$

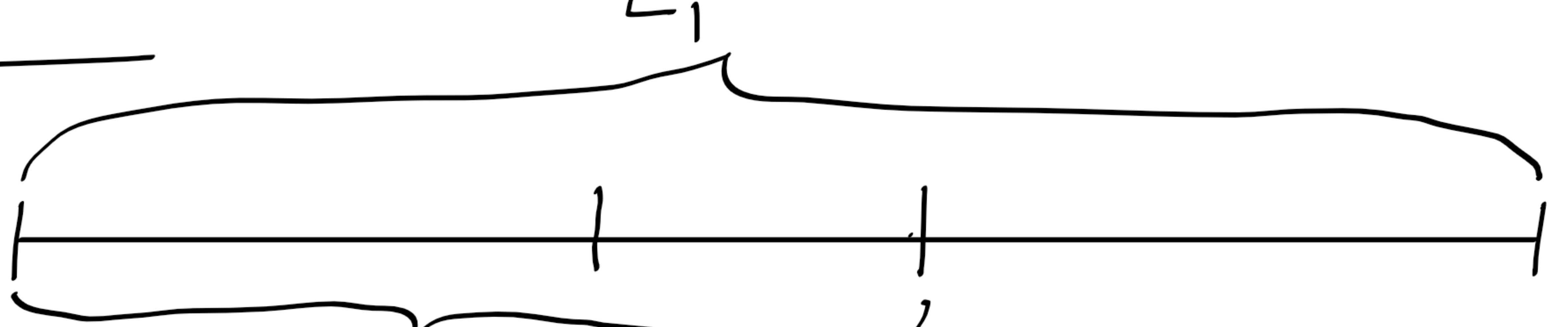
N even

$$\approx 0.71$$

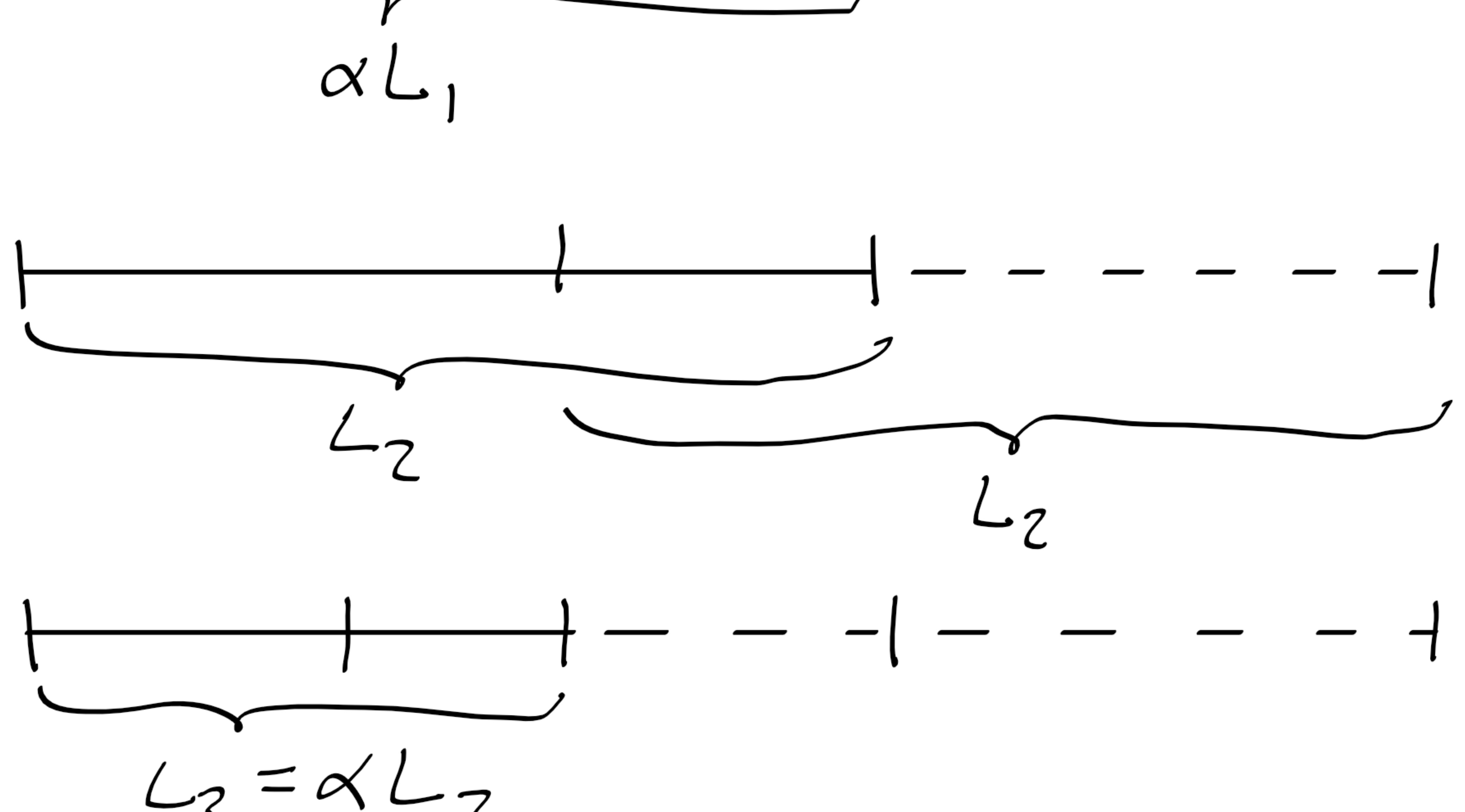


Golden Section Search

Start (one function eval.)



One new function evaluation per iteration



Symmetry gives

$$\alpha L_2 = L_1 - L_2 \iff$$

$$\alpha^2 L_1 = L_1 - \alpha L_1 \iff$$

$$\alpha^2 = 1 - \alpha \iff$$

$$\alpha^2 + \alpha - 1 = 0 \iff \alpha = -\frac{1}{2} \pm \sqrt{\frac{1}{4} + 1} = \frac{-1 + \sqrt{5}}{2} \approx 0.618$$

Thus $L_N = \alpha^{N-1} L_1$

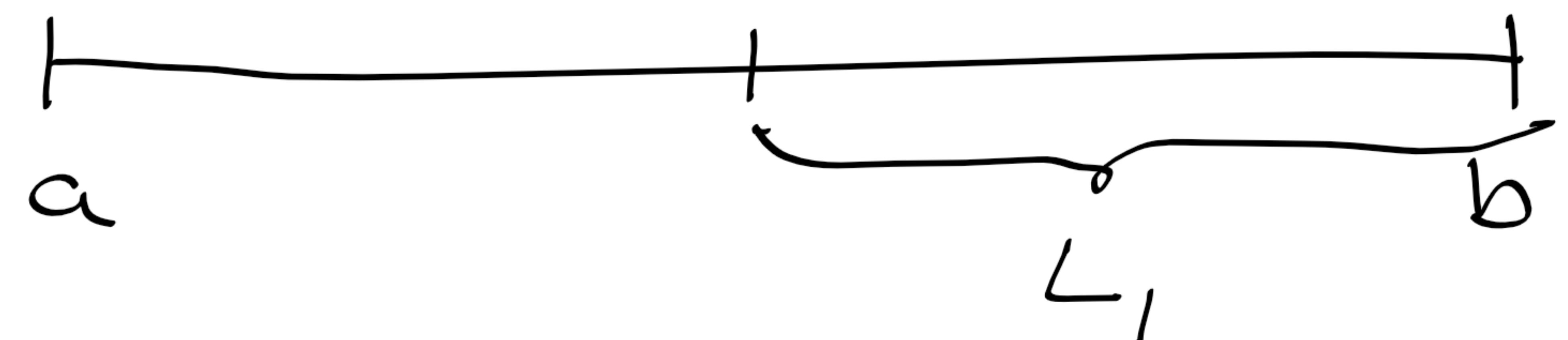
Bisection = check $F'(\frac{a+b}{2})$

$$L_0 = b - a$$

$$L_1 = \frac{1}{2} L_0$$

:

$$L_N = \frac{1}{2^N} L_0$$



Armijo's rule

λ is not too big
if $F(\lambda) \leq T(\lambda)$

λ is not too small

if $F(\alpha\lambda) \geq T(\alpha\lambda)$

$\alpha = 2$ typical

