Variational Methods & Optimal Control

lecture 02

Matthew Roughan
<matthew.roughan@adelaide.edu.au>

Discipline of Applied Mathematics
School of Mathematical Sciences
University of Adelaide

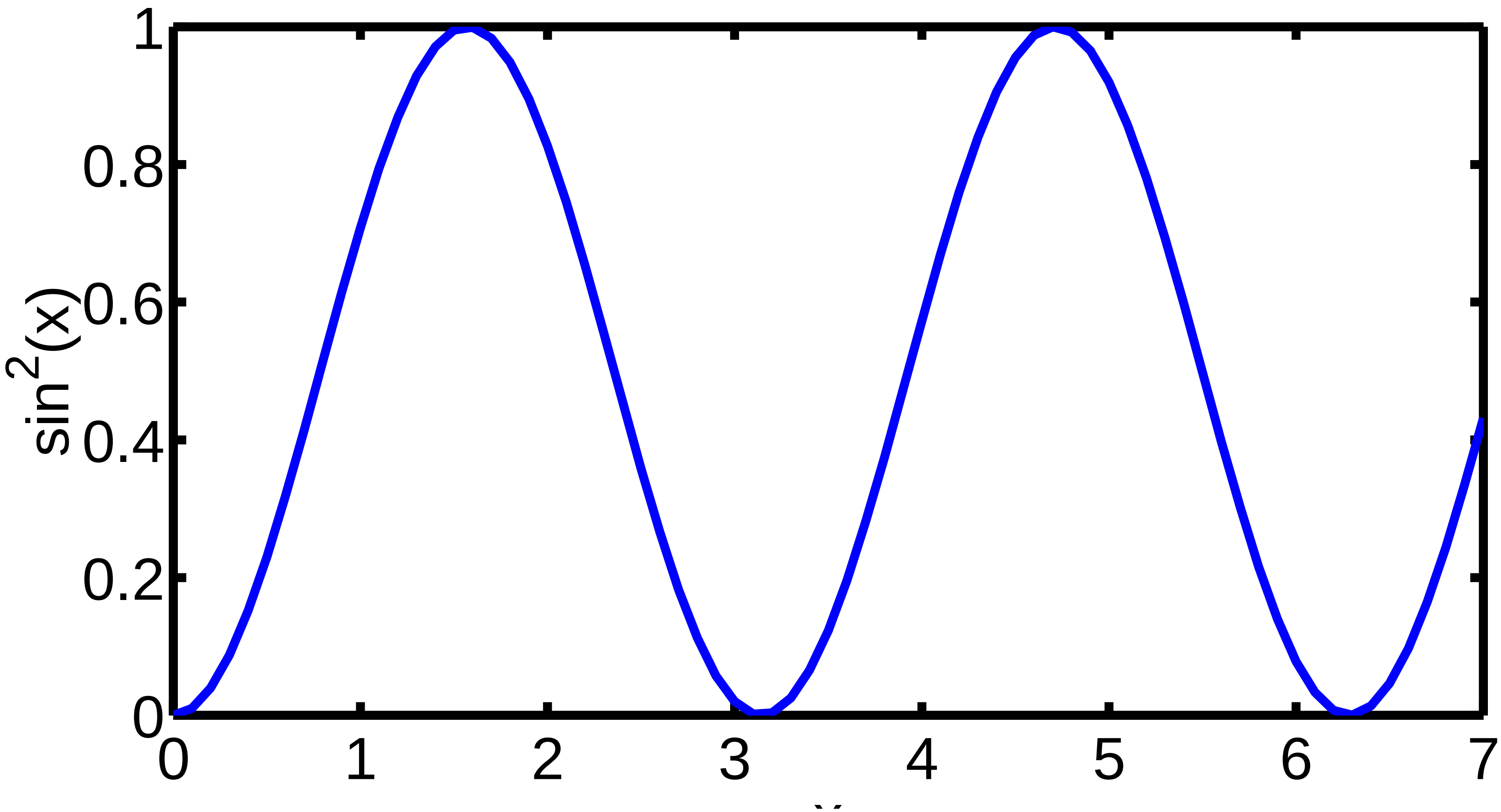
April 14, 2016

Revision, part ii

Extrema of functions of multiple variables. Taylor's theorem and the chain rule in N-D. Hessians and classification of extrema.

Extrema of functions of one variable

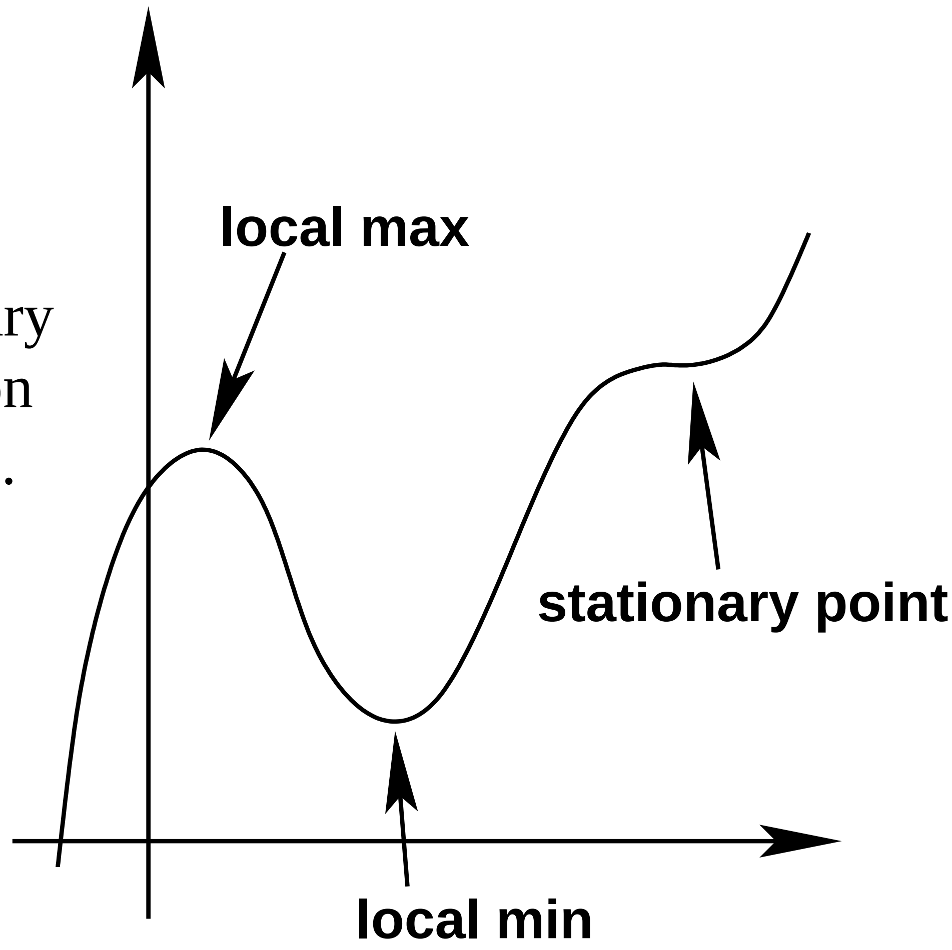
Local extrema have $f'(x) = 0$
includes maxima, minima, and stationary points of inflection



Classification of extrema

Local extrema have $f'(x) = 0$

- $f''(x) > 0$ local minima
- $f''(x) < 0$ local maxima
- $f''(x) = 0$ it might be a stationary point of inflection, depending on higher order derivatives, e.g. x^4.



Functions of n variables

- Let Ω be a closed region of $\mathbb{R}^n$, i.e. $\Omega \subset \mathbb{R}^n$
- Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \Omega$
- Let $f : \Omega \to \mathbb{R}$
- A local minima if $f(\cdot)$ is point $\mathbf{x}$ such that there exists $\delta > 0$ where

$$f(\hat{\mathbf{x}}) \geq f(\mathbf{x})$$

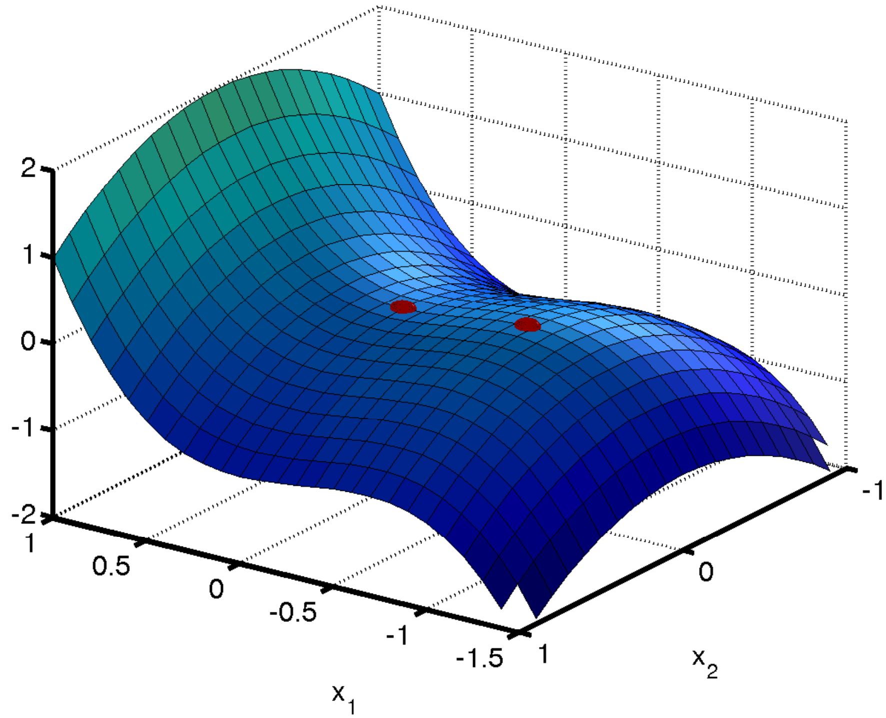
  for any $\hat{\mathbf{x}} \in B(\mathbf{x}; \delta)$.
- A global minima of $f(\cdot)$ on Ω is point $\mathbf{x}$ such that

$$f(\hat{\mathbf{x}}) \geq f(\mathbf{x})$$

  for any $\hat{\mathbf{x}} \in \Omega$.

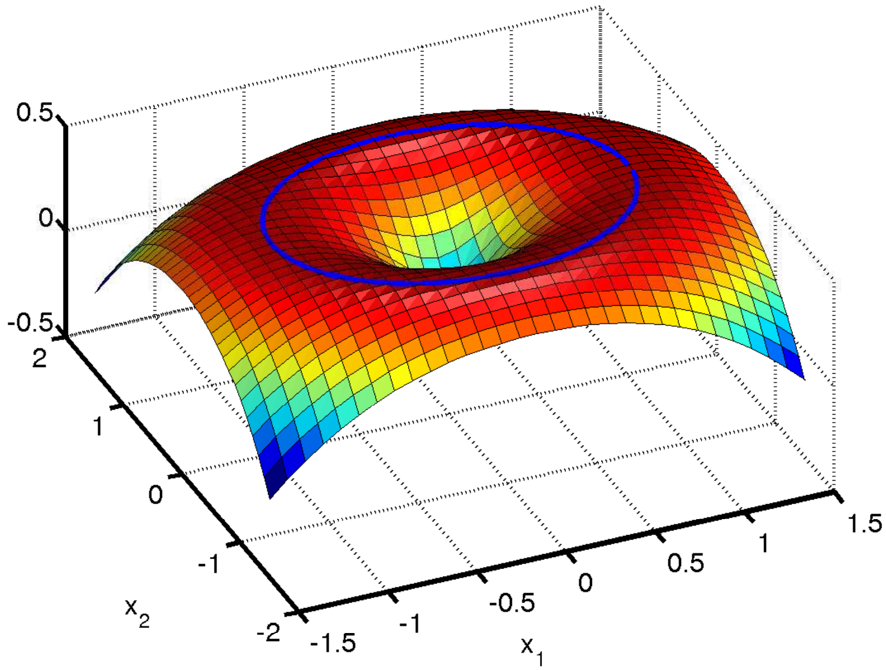
2D example 1

- $f(x_1, x_2) = x_1^2 - x_2^2 + x_1^3$
- local maximum at $(-2/3, 0)$
- saddle point at $(0, 0)$



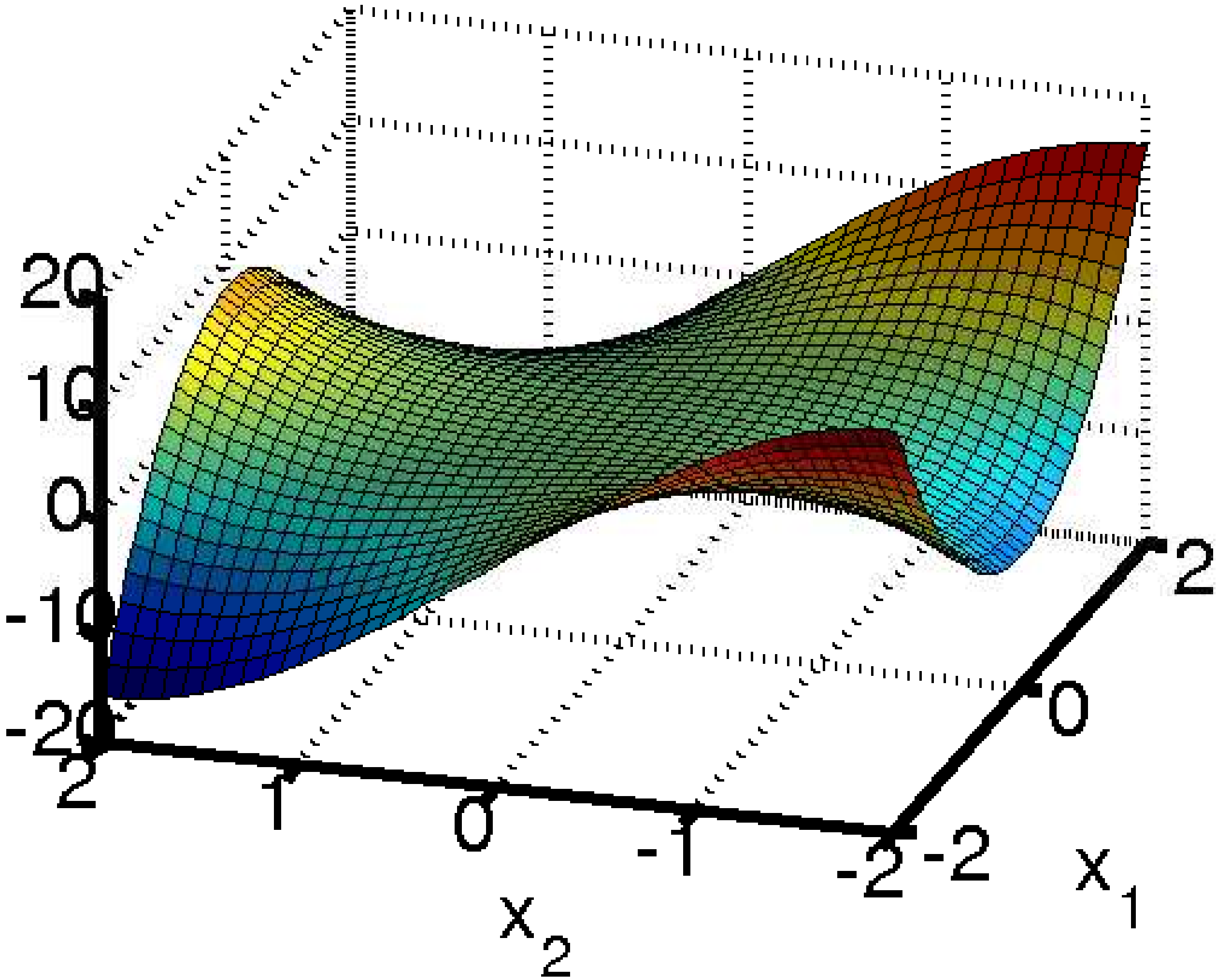
2D example 2

- $f(x_1, x_2) = r - 1/2r^2$, where $r = \sqrt{x_1^2 + x_2^2}$
- global maxima on curve $r = 1$
- local minima at $r = 0$



2D example 3

- $f(x_1, x_2) = x_2^3 - 3x_1^2 x_2$
- Monkey saddle at $(0,0)$



The Chain rule

The derivative of a function $f(x_1, x_2)$ along a line described parametrically by $(x_1(t), x_2(t))$

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt}$$

Another way to think of this is as the directional derivative formed from the dot product of grad and the direction of the line, e.g.,

$$\begin{aligned} \frac{df}{dt} &= \nabla f \cdot \frac{d\mathbf{x}}{dt} \\ &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) \cdot \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}\right) \end{aligned}$$

Chain Rule for more variables

The chain rule (for a function of more than one variable) $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$, where we want to find the derivative of a function $f(\mathbf{x})$ along a line described parametrically by $(x_1(t), x_2(t), \ldots, x_n(t))$ then we take

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}$$

or alternatively

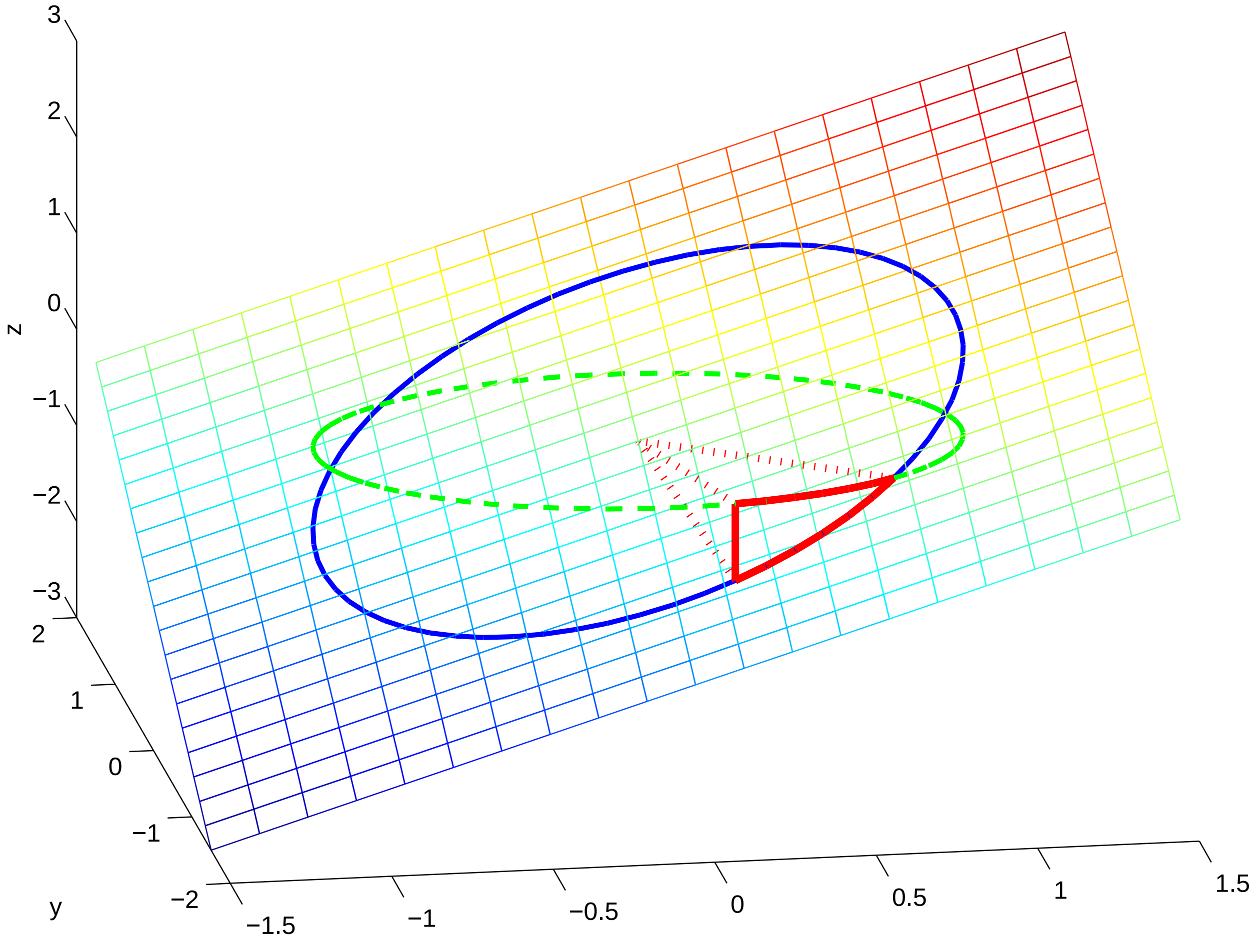
$$\begin{aligned}\frac{df}{dt} &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) \bullet \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_n}{dt}\right) \\ &= \nabla f \bullet \frac{d\mathbf{x}}{dt}\end{aligned}$$

A graphical example

For a function of two variables $f(x,y)$ we get

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

$$\begin{aligned} f(x,y) &= x+y \\ x &= \cos t \\ y &= \sin t \end{aligned}$$

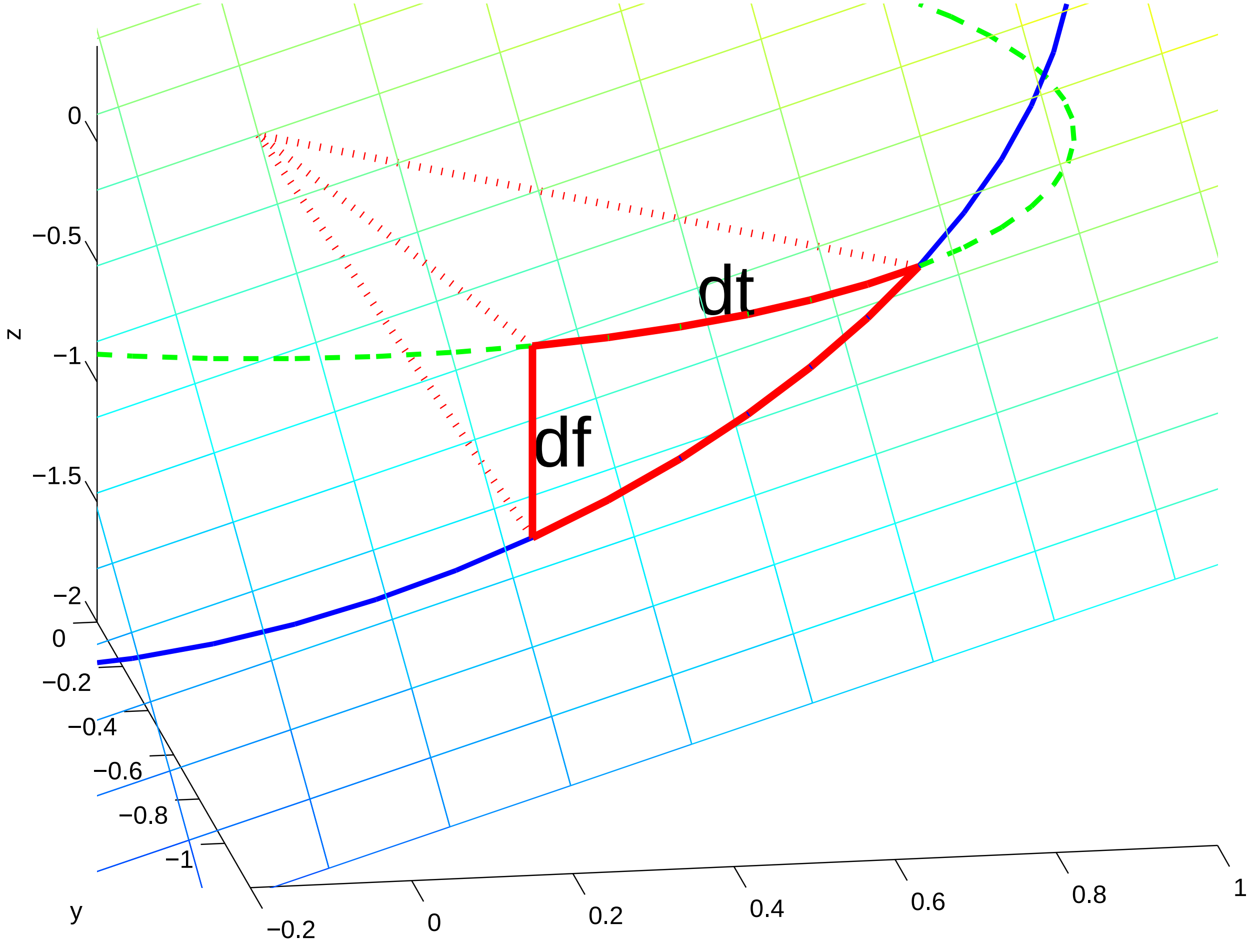


A graphical example

For a function of two variables $f(x,y)$ we get

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

$$\begin{aligned} f(x,y) &= x+y \\ x &= \cos t \\ y &= \sin t \end{aligned}$$

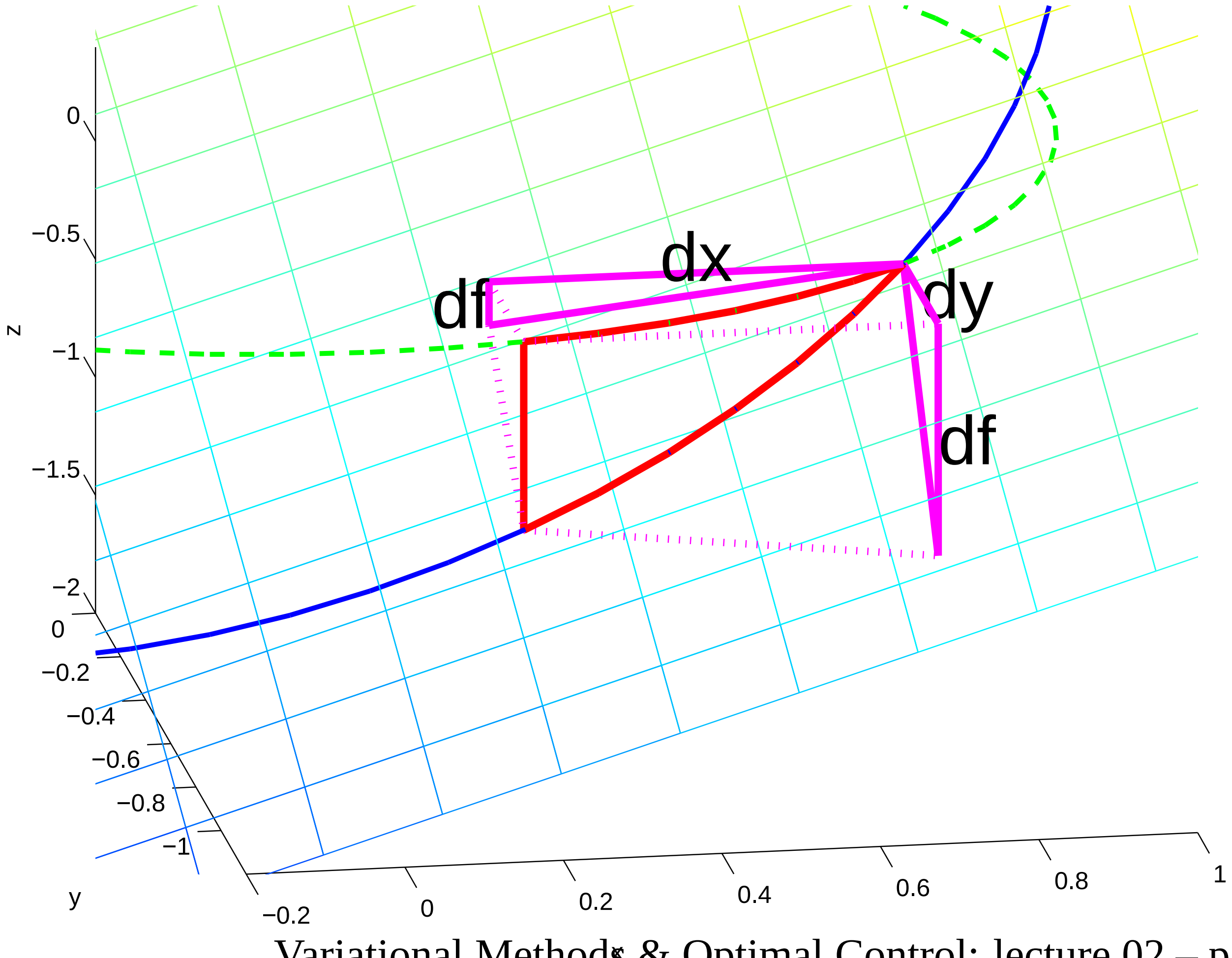


A graphical example

For a function of two variables $f(x,y)$ we get

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

$$\begin{aligned} f(x,y) &= x+y \\ x &= \cos t \\ y &= \sin t \end{aligned}$$



Chain Rule Derivation

By the definition

$$\frac{df}{dt} = \lim_{\varepsilon \to 0} \frac{f(x(t+\varepsilon), y(t+\varepsilon)) - f(x(t), y(t))}{\varepsilon}$$

But note that from Taylor's theorem

$$x(t+\varepsilon) = x(t) + \varepsilon x'(t) + O(\varepsilon^2).$$

As we consider the limit as $\varepsilon \to 0$ we may ignore the $O(\varepsilon^2)$ term, to get

$$\frac{df}{dt} = \lim_{\varepsilon \to 0} \frac{f(x(t) + \varepsilon x'(t), y(t) + \varepsilon y'(t)) - f(x(t), y(t))}{\varepsilon}$$

Chain Rule Derivation

$$
\begin{aligned}
\frac{df}{dt} &= \lim_{\varepsilon\to 0}\frac{f(x(t)+\varepsilon x'(t),y(t)+\varepsilon y'(t))-f(x(t),y(t))}{\varepsilon}\\
&= \lim_{\varepsilon\to 0}\frac{f(x(t)+\varepsilon x'(t),y(t)+\varepsilon y'(t))-f(x(t),y(t)+\varepsilon y'(t))}{\varepsilon}\\
&\qquad +\lim_{\varepsilon\to 0}\frac{f(x(t),y(t)+\varepsilon y'(t))-f(x(t),y(t))}{\varepsilon}\\
&= x'(t)\lim_{\varepsilon\to 0}\frac{f(x(t)+\varepsilon x'(t),y(t)+\varepsilon y'(t))-f(x(t),y(t)+\varepsilon y'(t))}{\varepsilon x'(t)}\\
&\qquad +y'(t)\lim_{\varepsilon\to 0}\frac{f(x(t),y(t)+\varepsilon y'(t))-f(x(t),y(t))}{\varepsilon y'(t)}
\end{aligned}
$$

Chain Rule Derivation

$$
\begin{aligned}
\frac{df}{dt} &= x'(t)\lim_{\varepsilon\to 0}\frac{f(x(t)+\varepsilon x'(t),y(t)+\varepsilon y'(t)) - f(x(t),y(t)+\varepsilon y'(t))}{\varepsilon x'(t)} \\
&\quad + y'(t)\lim_{\varepsilon\to 0}\frac{f(x(t),y(t)+\varepsilon y'(t)) - f(x(t),y(t))}{\varepsilon y'(t)} \\
&= x'(t)\lim_{\varepsilon_x\to 0}\frac{f(x(t)+\varepsilon_x,y(t)+\varepsilon y'(t)) - f(x(t),y(t)+\varepsilon y'(t))}{\varepsilon_x} \\
&\quad + y'(t)\lim_{\varepsilon_y\to 0}\frac{f(x(t),y(t)+\varepsilon_y) - f(x(t),y(t))}{\varepsilon_y} \\
&= x'(t)\frac{\partial f}{\partial x} + y'(t)\frac{\partial f}{\partial y}
\end{aligned}
$$

which is the chain rule!

Chain Rule special case

When we only have one variable, we simple want to calculate the derivative of a function f of another function x, e.g.

$$\frac{d}{dt} f(x(t)) = \frac{df}{dx}\frac{dx}{dt}$$

Another way of writing this is

$$\frac{d}{dt} f(x(t)) = f'\left[x(t)\right] x'(t),$$

which is the form you learnt in 1st year.

Taylor's theorem in 2D

$$f(x_1+\delta x_1, x_2+\delta x_2) = f(x_1,x_2) + \delta x_1 \frac{\partial f}{\partial x_1} + \delta x_2 \frac{\partial f}{\partial x_2}$$
$$+\frac{1}{2}\left[\delta x_1^2 \frac{\partial^2 f}{\partial x_1^2} + 2\delta x_1 \delta x_2 \frac{\partial^2 f}{\partial x_1 \partial x_2} + \delta x_2^2 \frac{\partial^2 f}{\partial x_2^2}\right] + \cdots$$

Write $(\delta x_1, \delta x_2) = \varepsilon \times (\eta_1, \eta_2)$

$$f(\mathbf{x}+\varepsilon\eta) = f(\mathbf{x}) + \varepsilon\left(\eta_1 \frac{\partial f}{\partial x_1} + \eta_2 \frac{\partial f}{\partial x_2}\right)$$
$$+\frac{\varepsilon^2}{2}\left[\eta_1^2 \frac{\partial^2 f}{\partial x_1^2} + 2\eta_1\eta_2 \frac{\partial^2 f}{\partial x_1 \partial x_2} + \eta_2^2 \frac{\partial^2 f}{\partial x_2^2}\right] + O(\varepsilon^3)$$

Taylor's theorem in N-D

$$f(\mathbf{x}+\delta\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{n} \delta x_i \frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i x_j} \delta x_i \delta x_j + O(\delta\mathbf{x}^3)$$

$$f(\mathbf{x}+\delta\mathbf{x}) = f(\mathbf{x}) + \delta\mathbf{x}^T \nabla f(\mathbf{x}) + \frac{1}{2}\delta\mathbf{x}^T H(\mathbf{x})\delta\mathbf{x} + O(\delta\mathbf{x}^3)$$

Where $H(\mathbf{x})$ is the Hessian matrix

$$H(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Maxima of N variables

If a smooth function $f(\mathbf{x})$ has a local extrema at $\mathbf{x}$ then

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)^T = \mathbf{0}$$

A sufficient condition for the extrema $\mathbf{x}$ to be a local minimum is for the quadratic form

$$Q(\delta x_1, \ldots, \delta x_n) = \delta\mathbf{x}^T H(\mathbf{x})\delta\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\delta x_i \delta x_j$$

to be strictly positive definite.

Quadratic forms

A quadratic form

$$Q(\mathbf{x}) = \sum_{i,j} a_{ij} x_i x_j = \mathbf{x}^T A \mathbf{x}$$

is said to be positive definite if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

A quadratic form is positive definite iff every eigenvalue of A is greater than zero.

A quadratic form is positive definite if all the principal minors in the top-left corner of A are positive, in other words

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \cdots$$

Notes on maxima and minima

- maxima of $f(x)$ are minima of $-f(x)$.
- haven't said anything about non-differentiable functions
- if continuous in the interval, must achieve maximum (minimum) in the interval

Calculus of variations

- We are not maximizing the value of a function.
- We are maximizing a **functional**
  — a function of a function
- Can think of it as an ∞-dimensional max. problem.
  - can choose between different functions
  - function sits in ∞-dimensional vector space
- This might take some effort.

Functionals

A **Functional** maps an element of a vector space (e.g. a space containing functions) to a real number, e.g. $F : S \rightarrow \mathbb{R}$.

Example Functionals

$$
\begin{aligned}
F\{y(x)\} &= |y(0)| \\
F\{y(x)\} &= \max_x \{y(x)\} \\
F\{y(x)\} &= \left.\frac{dy}{dx}\right|_{x=1} \\
F\{y(x)\} &= y(0) + y(1) \\
F\{y(x)\} &= \sum_{n=0}^{N} a_n y(n)
\end{aligned}
$$

Integral functionals

- Previous functionals not very interesting.
- Easy to find $y(x)$ which minimizes these.
- Integral functionals are more interesting.
- Example integral functionals

$$
\begin{aligned}
F\{y\} &= \int_a^b y(x)dx \\
F\{y\} &= \int_a^b f(x)y(x)dx \\
F\{y\} &= \int_a^b \sqrt{1+\left(\frac{dy}{dx}\right)^2}\, dx
\end{aligned}
$$

Some simple cases

$$\begin{aligned} F\{y\} &= \int_{a(\varepsilon)}^{b(\varepsilon)} y(x,\varepsilon)dx \\ \frac{dF}{d\varepsilon} &= y(b,\varepsilon)\frac{db}{d\varepsilon} - y(a,\varepsilon)\frac{da}{d\varepsilon} + \int_{a(\varepsilon)}^{b(\varepsilon)} \frac{\partial y(x,\varepsilon)}{\partial \varepsilon}dx \end{aligned}$$

If a and b are fixed then

$$\begin{aligned} \frac{da}{d\varepsilon} &= 0 \\ \frac{db}{d\varepsilon} &= 0 \end{aligned}$$

and so the derivative of the integral becomes the integral of the derivative.

Crude Brachystochrone

Brachystochrone involves the functional

$$F\{y\} = \int_{x_0}^{x_1} \sqrt{\frac{1+y'^2}{y}}\,dx$$

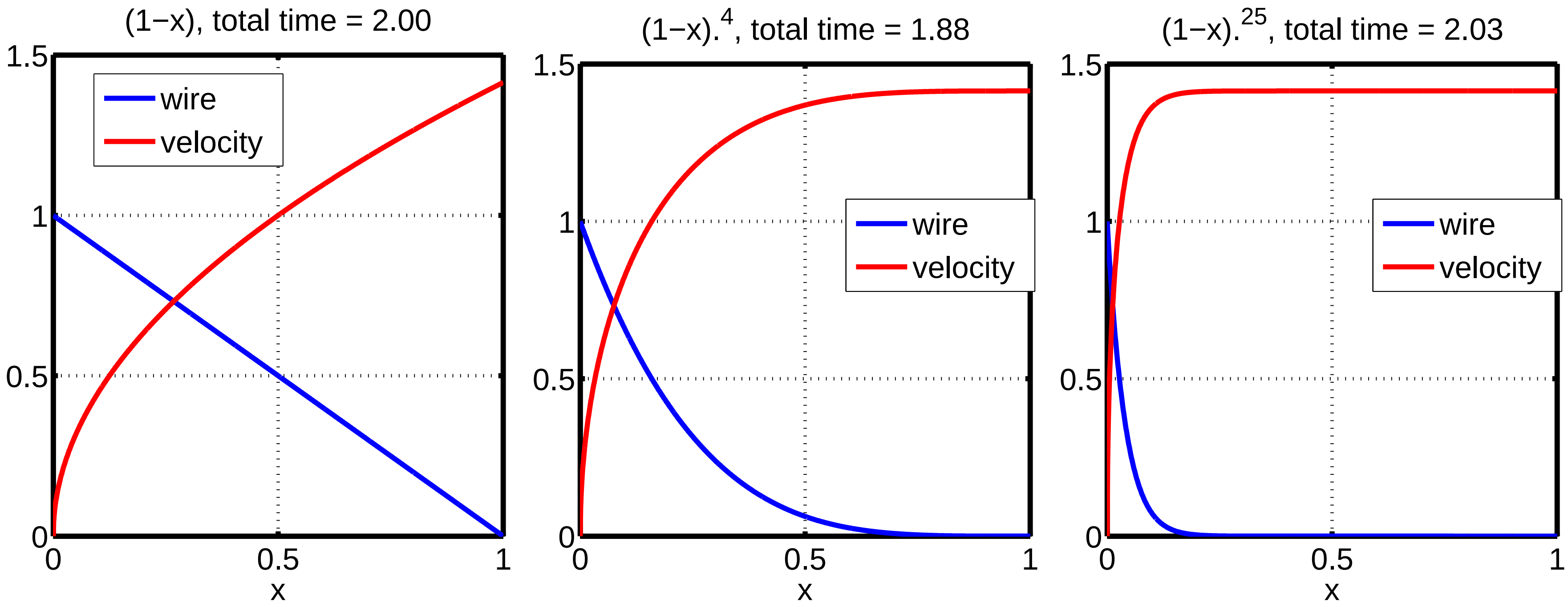
Let us guess that the brachystochrone takes the form

$$y(x,\varepsilon) = (1-x)^{\varepsilon}$$

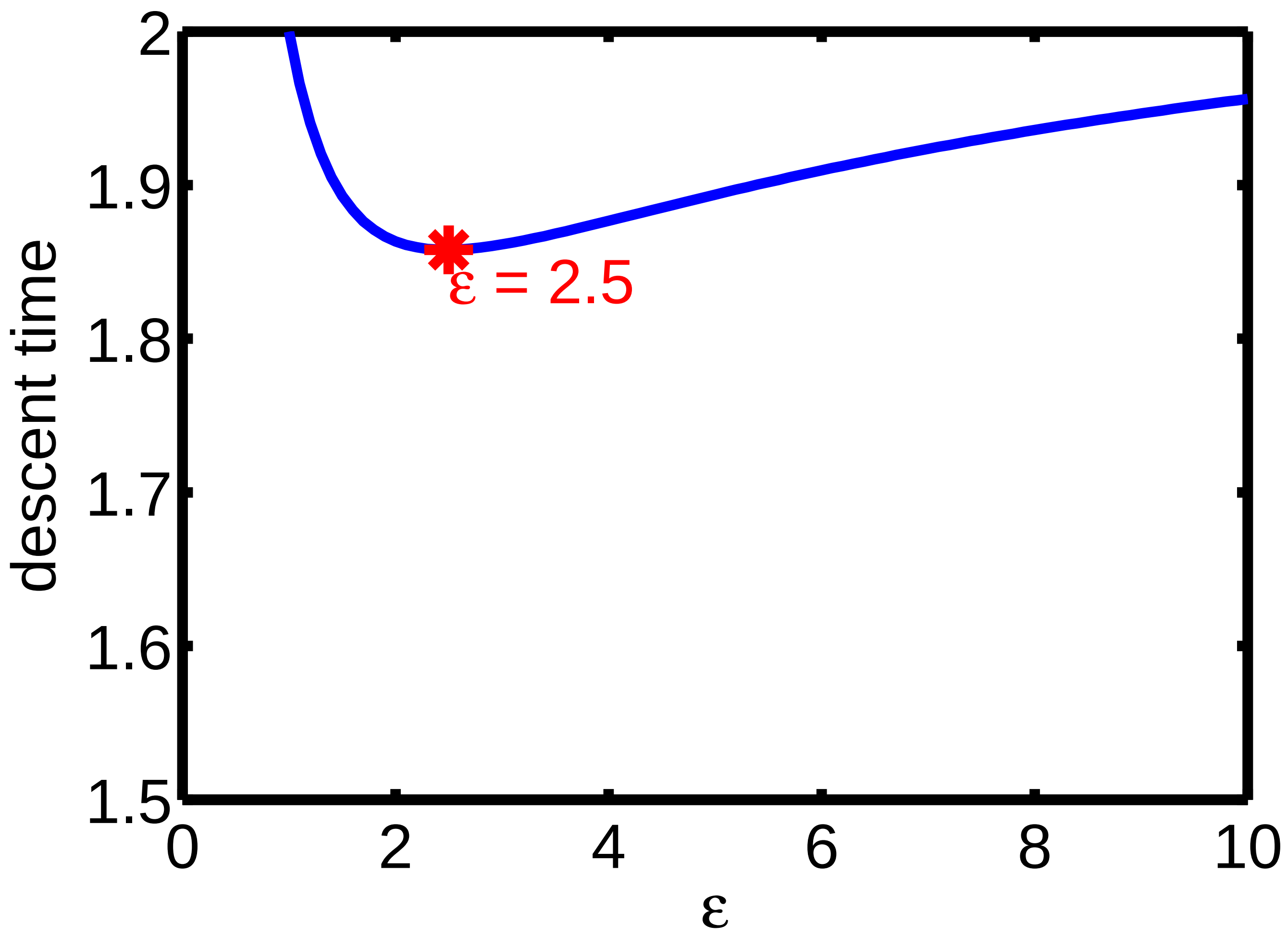
We could calculate the derivative WRT ε as above and compute the stationary points by finding

$$\frac{dF}{d\varepsilon} = 0$$

Crude Brachystochrone



Crude Brachystochrone



- but what if the family of curves doesn't contain the maximum?

Extra bits

Notation

- $f(\mathbf{x}) : S \to \mathbb{R}$ denotes a function that maps the set $S \subset \mathbb{R}^n$ to a real number.
- $\frac{\partial^n f}{\partial x_i^n}$ denotes the nth partial derivative of $f(\mathbf{x})$, with respect to x_i.
- the ε-neighborhood under the Euclidean norm is $B(\mathbf{x};\varepsilon) = \{\hat{\mathbf{x}} \in \mathbb{R}^n \mid \|\hat{\mathbf{x}} - \mathbf{x}\|_2 < \varepsilon\}$
- The Euclidean norm in $\mathbb{R}^n$ is $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$
- $F\{y\}$ denotes a functional of the function $y(x)$.