Optimization background

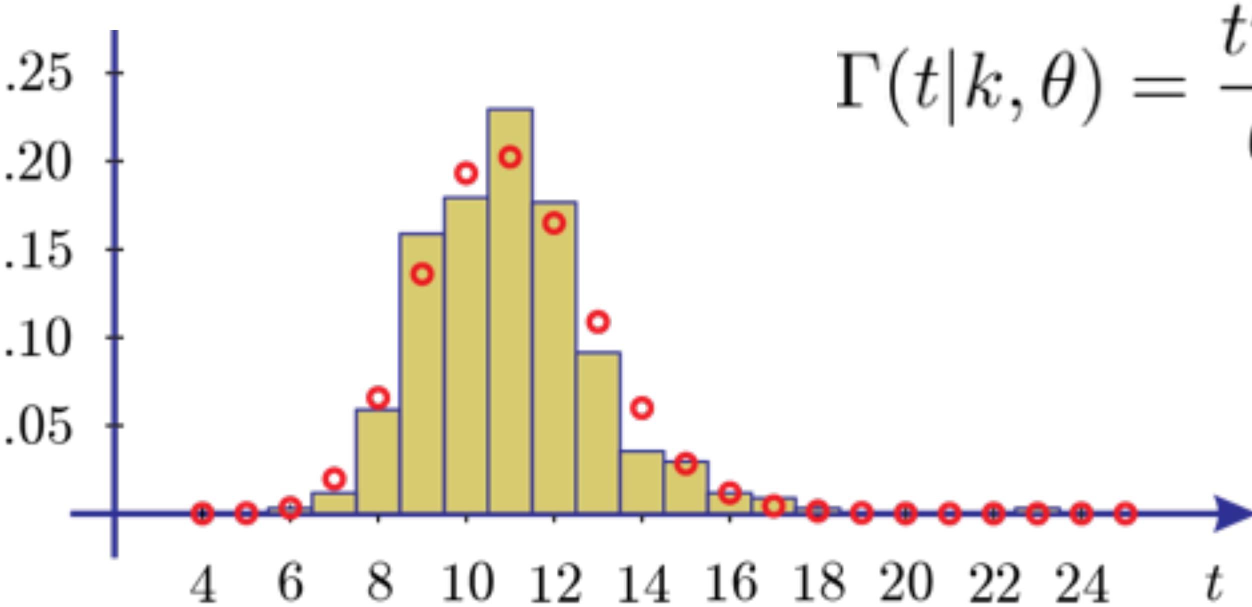
Reminders

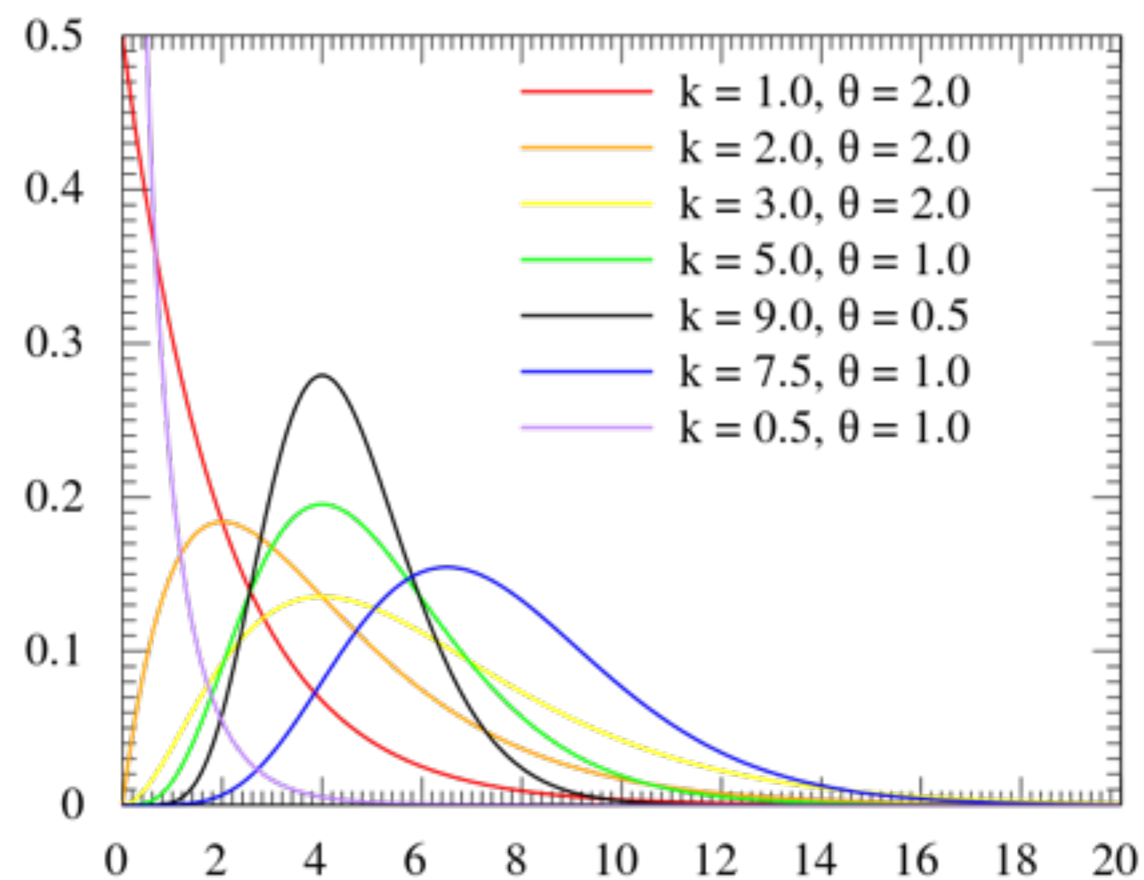
- Assignment #1 is due this Wednesday
  - For the last question, be clear about your variables and specify an explicit optimization; you do not have to solve the optimization
- Next readings: Chapters 4 and 5, Section 6.1
  - Chapter 4: Linear regression
  - Chapter 5: Generalized linear models
  - Section 6.1 Beginning of classification, with logistic regression
- Office hours

Exercise

- We modeled commute times as p(x l theta)
- How might we predict the commute time for today?

$$\Gamma(t|k,\theta) = \frac{t^{k-1}e^{-\frac{t}{\theta}}}{\theta^k \Gamma(k)}$$





Thought questions

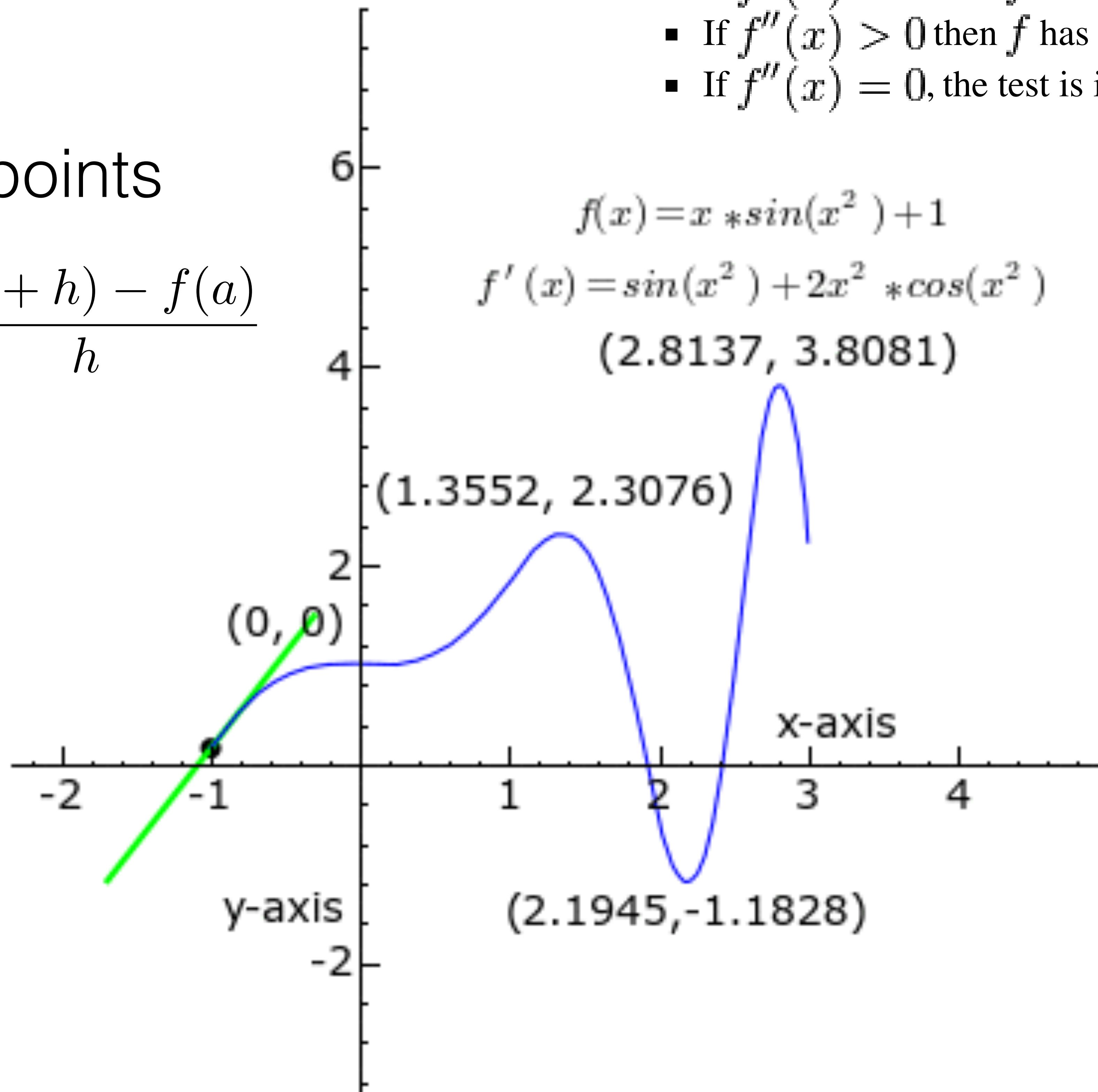
- It's stated that MAP and ML converge to the same solution for large data sets. Is there a difference in the rates of convergence between these two approaches, i.e. could it potentially be more useful to use MAP rather than ML despite being harder to calculate if MAP had a faster rate of convergence?
  - Imagine an oracle gave you the true parameter before learning
  - Intuitively if the prior chosen for MAP heavily biases the parameter to be close to the true parameter, then we should converge (get to the optimal solution) faster, since we started closer to it
  - If log prior is strongly convex, can improve convergence rate
  - Different types of convergence: convergence to solution for fixed # of samples and convergence to true parameters with infinite samples

Univariate optimization

- Minima
- Maxima
- Saddle points

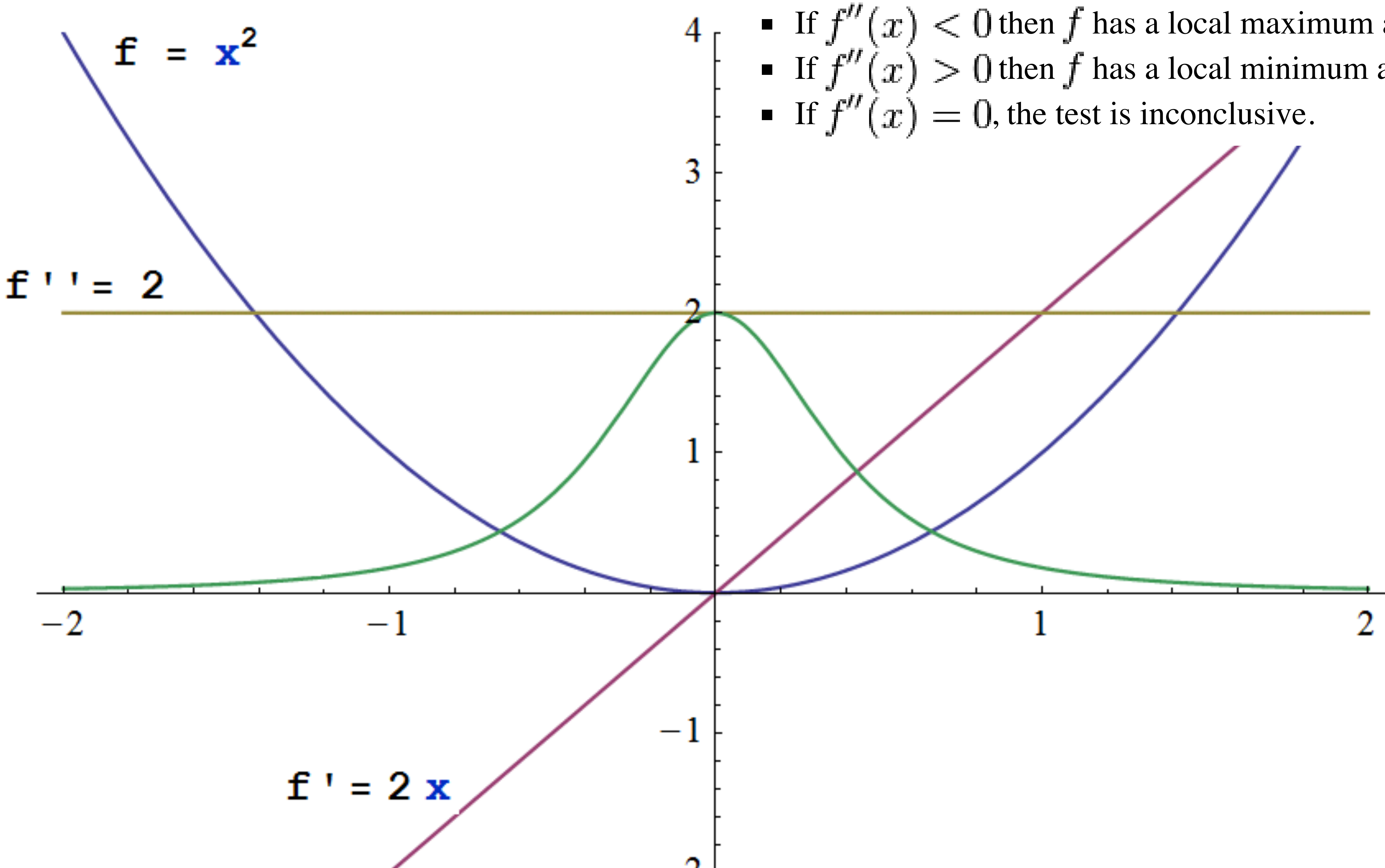
$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.



Second derivative test

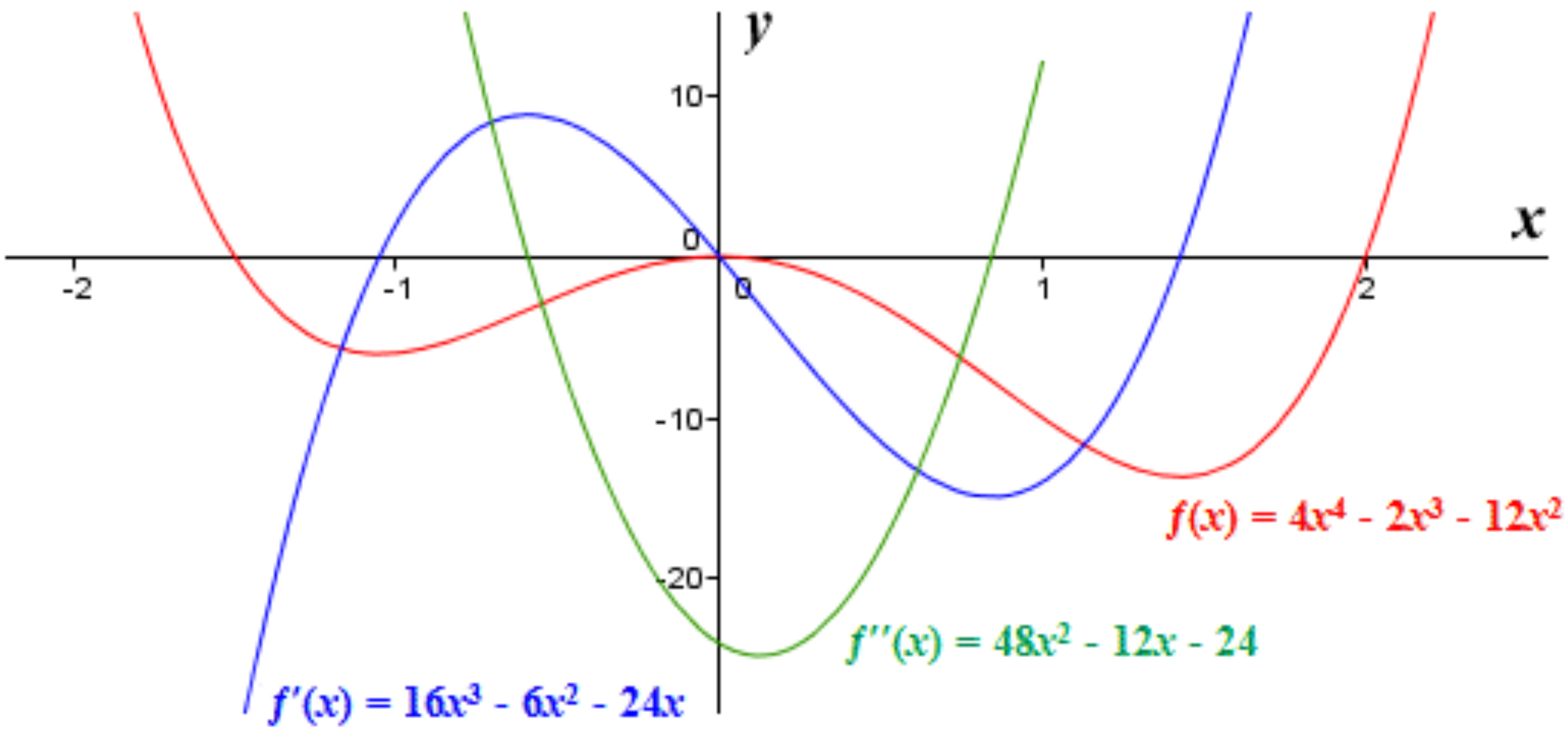
- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.



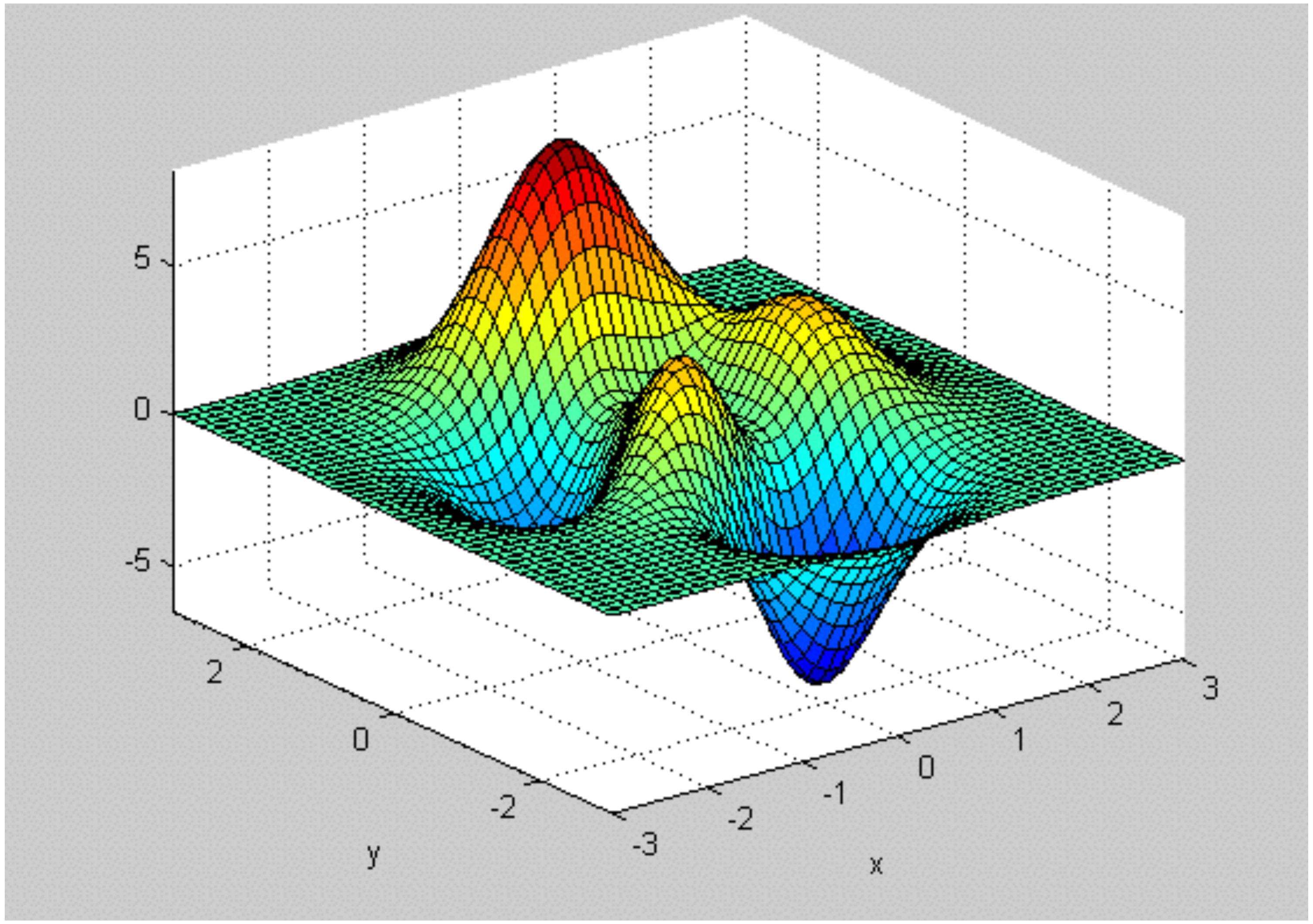
Ignore the green line

Second derivative test

- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.



Multivariate optimization or Multivariable optimization



First-order conditions

$$\nabla f = \left[\frac{\partial f}{\partial x_1}, \ \ldots, \ \frac{\partial f}{\partial x_d}\right]$$

- Gradient = 0 provides a minimum, maximum or saddle point
- Gradient gives direction of steepest ascent

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

Directional second derivative

$$\text{At stationary point } \mathbf{w}^*, \nabla f(\mathbf{w}) = \mathbf{0}$$

$$\mathbf{w}(t) = \mathbf{w}^* + t\mathbf{w}$$

$$g(t) = f(\mathbf{w}(t))$$

$$g'(0) = \nabla f(\mathbf{w}(t))^\top \mathbf{w} = 0$$

$$g''(0) = \mathbf{w}^\top \nabla^2 f(\mathbf{w}(t))^\top \mathbf{w}$$

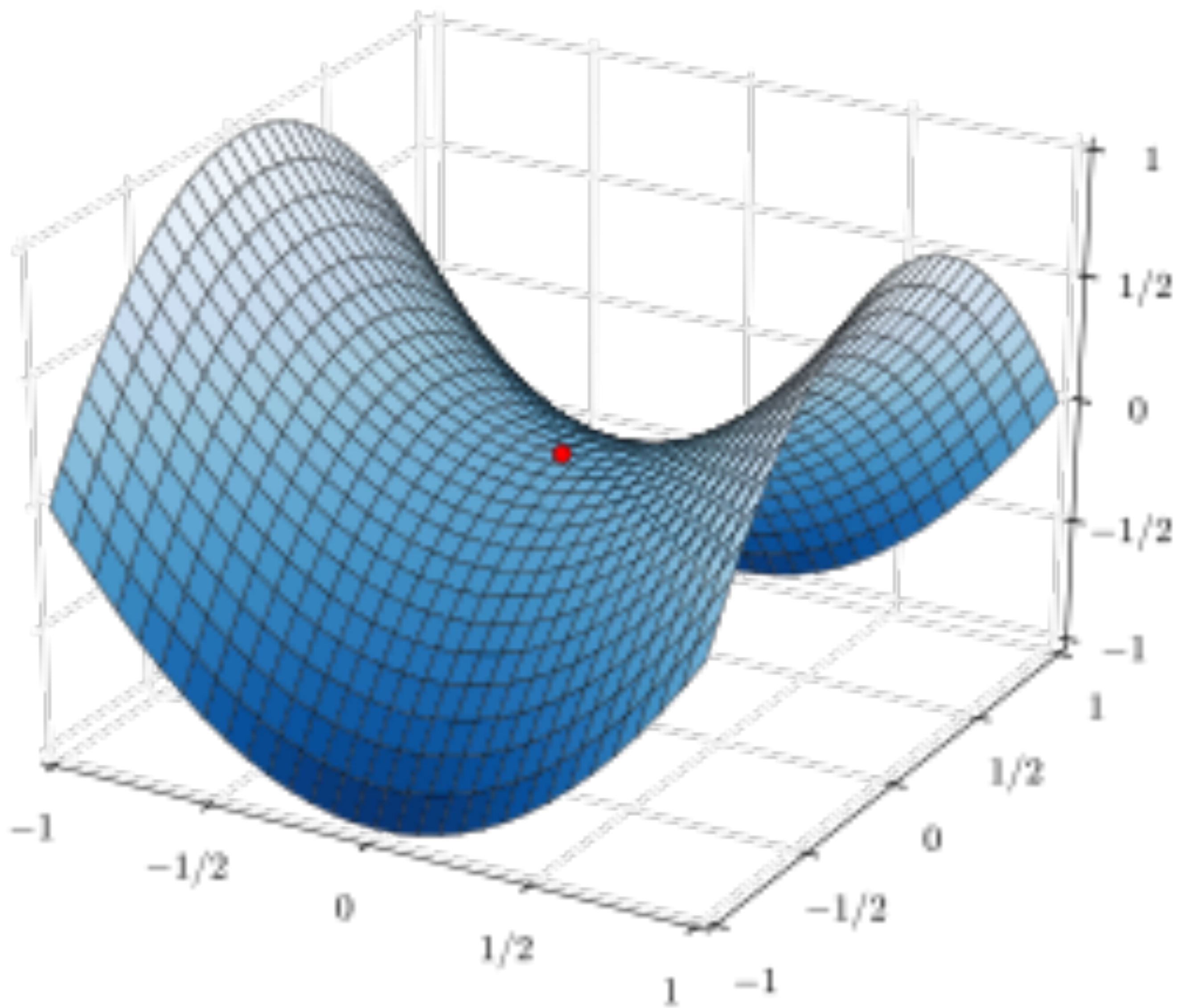
Intuition for second derivative test in univariate setting

$$0 < f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} = \lim_{h \to 0} \frac{f'(x+h) - 0}{h} = \lim_{h \to 0} \frac{f'(x+h)}{h}.$$

Thus, for h sufficiently small we get

$$\frac{f'(x+h)}{h} > 0$$

Hessian intuition



Hessian

- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.

$$g''(0) = \mathbf{w}^\top \nabla^2 f(\mathbf{w}(t))^\top \mathbf{w}$$

$$\nabla^2 f = \quad \mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Hessian

- If $f''(x) < 0$ then f has a local maximum at x.
- If $f''(x) > 0$ then f has a local minimum at x.
- If $f''(x) = 0$, the test is inconclusive.

$$g''(0) = \mathbf{w}^\top \nabla^2 f(\mathbf{w}(t))^\top \mathbf{w}$$

$$\text{Positive definite: } \mathbf{w}^\top \mathbf{H} \mathbf{w} > 0 \text{ for all } \mathbf{w} \neq \mathbf{0}$$

$$\text{Negative definite: } \mathbf{w}^\top \mathbf{H} \mathbf{w} < 0 \text{ for all } \mathbf{w} \neq \mathbf{0}$$

- If H is positive definite at x, then local minimum at x
- If H is negative definite at x, then local maximum at x
- If H has both positive and negative eigenvalues at x, then a saddle point at x

Whiteboard

- Maximum likelihood for Gaussian with theta = (mu, sigma)
- Is the solution a local minimum, maximum or saddlepoint?