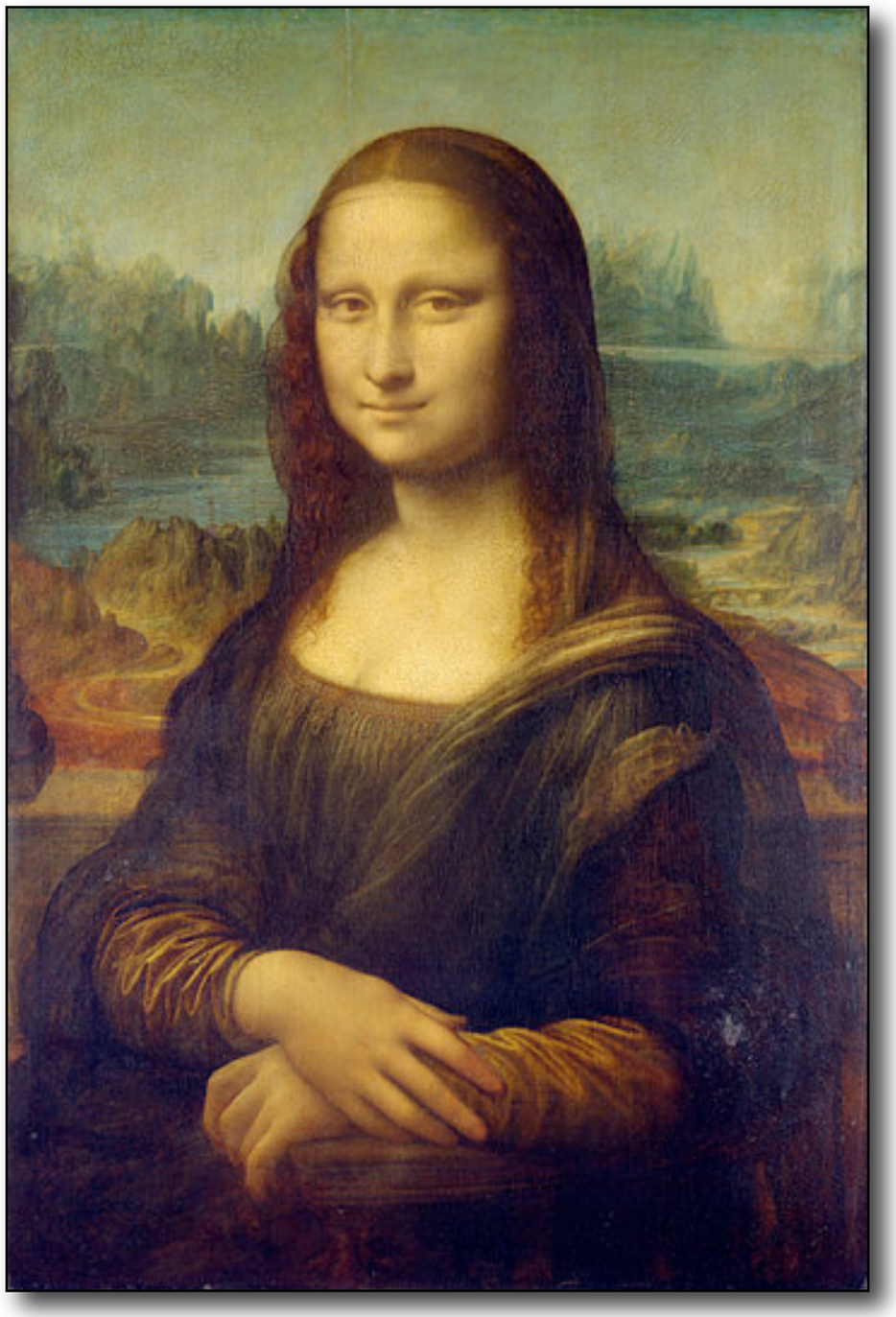
First-order methods
Convexity

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Administrivia

- Schedule posted:
  - Time for poster session: 3:30–6:30, Wed, Dec 12
  - Midterm: Tue, Nov 6 (in class)
  - HW1 will be released: hopefully Tue, Sep 4
- How to do scribing:
  - http://www.cs.cmu.edu/~aarti/Class/10704/lecs.html
- In case of mishaps with scribe signup sheet

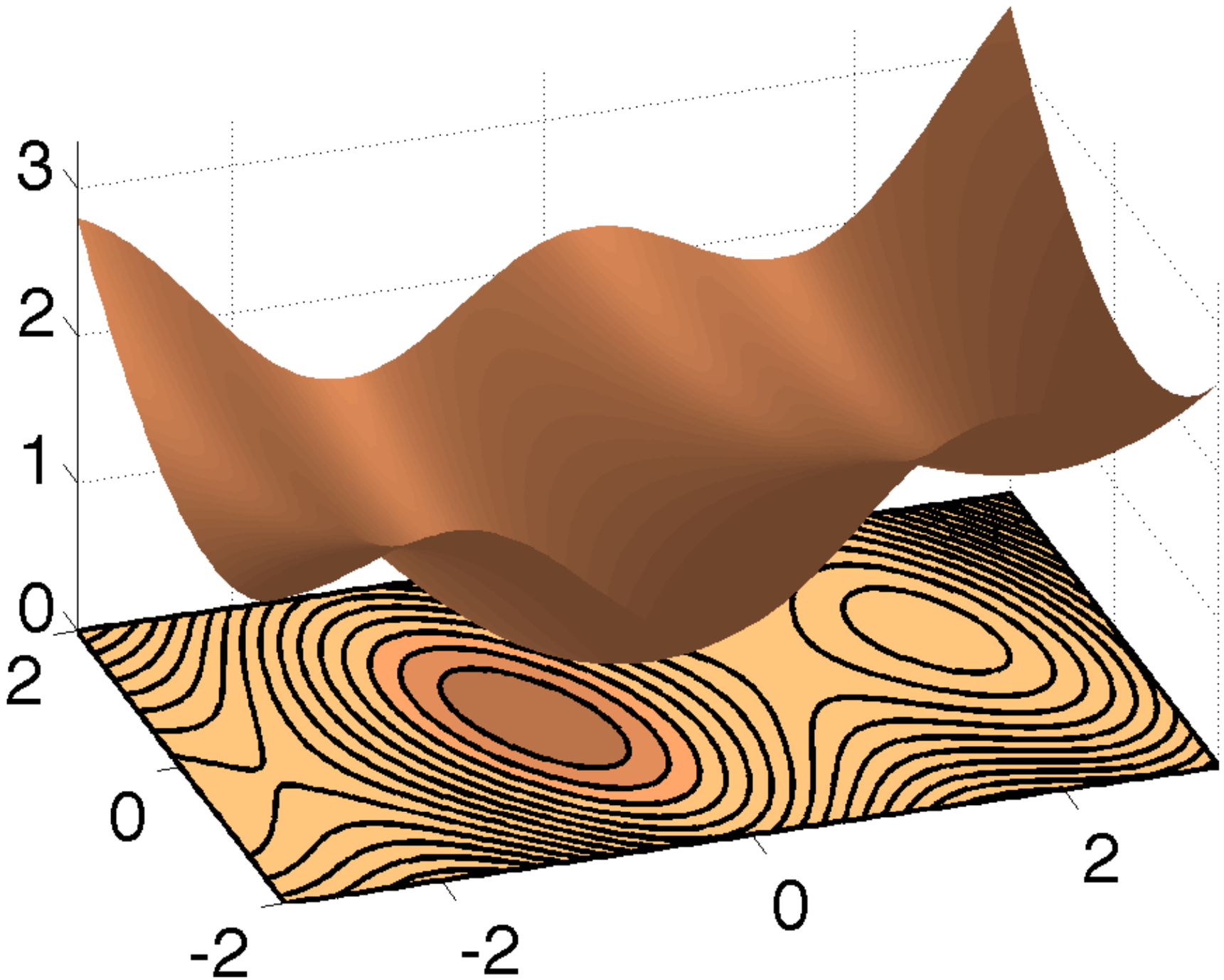
Worked ex: image understanding



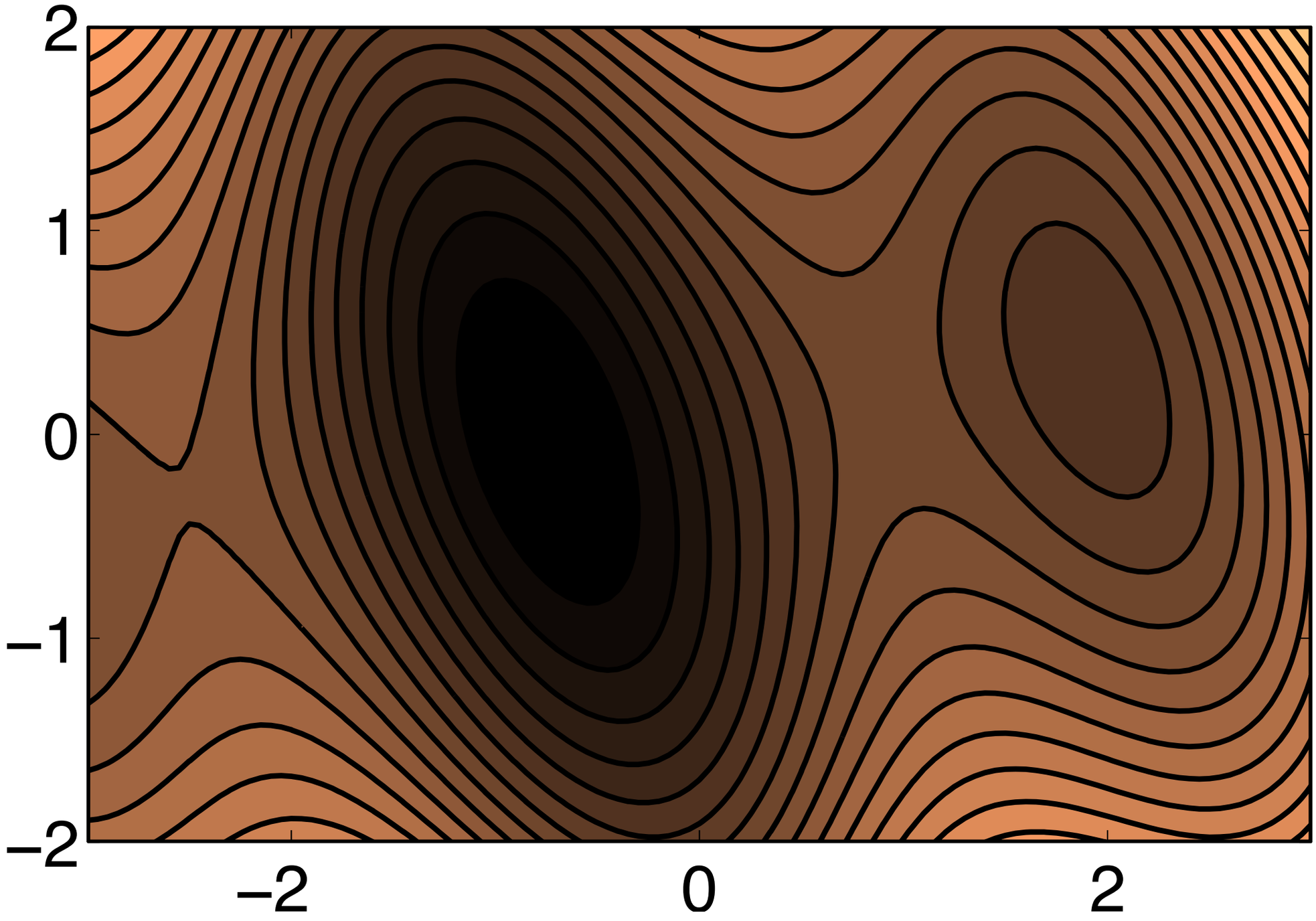
Gradient descent

- for k = 1, 2, …
  - $g_k \leftarrow \nabla f(x_k)$
  - $x_k \leftarrow x_{k-1} - t_k g_k$
- Choices: x_0, t_k, when to stop

Gradient descent: example



Gradient descent: example



In ML & stats

- Often have f(x) =
  - where i ~
- E.g., linear regression:
- Let:
  - then

When do we stop?

- ML/stats: held out data

- Early stopping
  -
  -

When do we stop?

- Using convergence bounds (see below)
  - usual form is:
  - need estimates of: