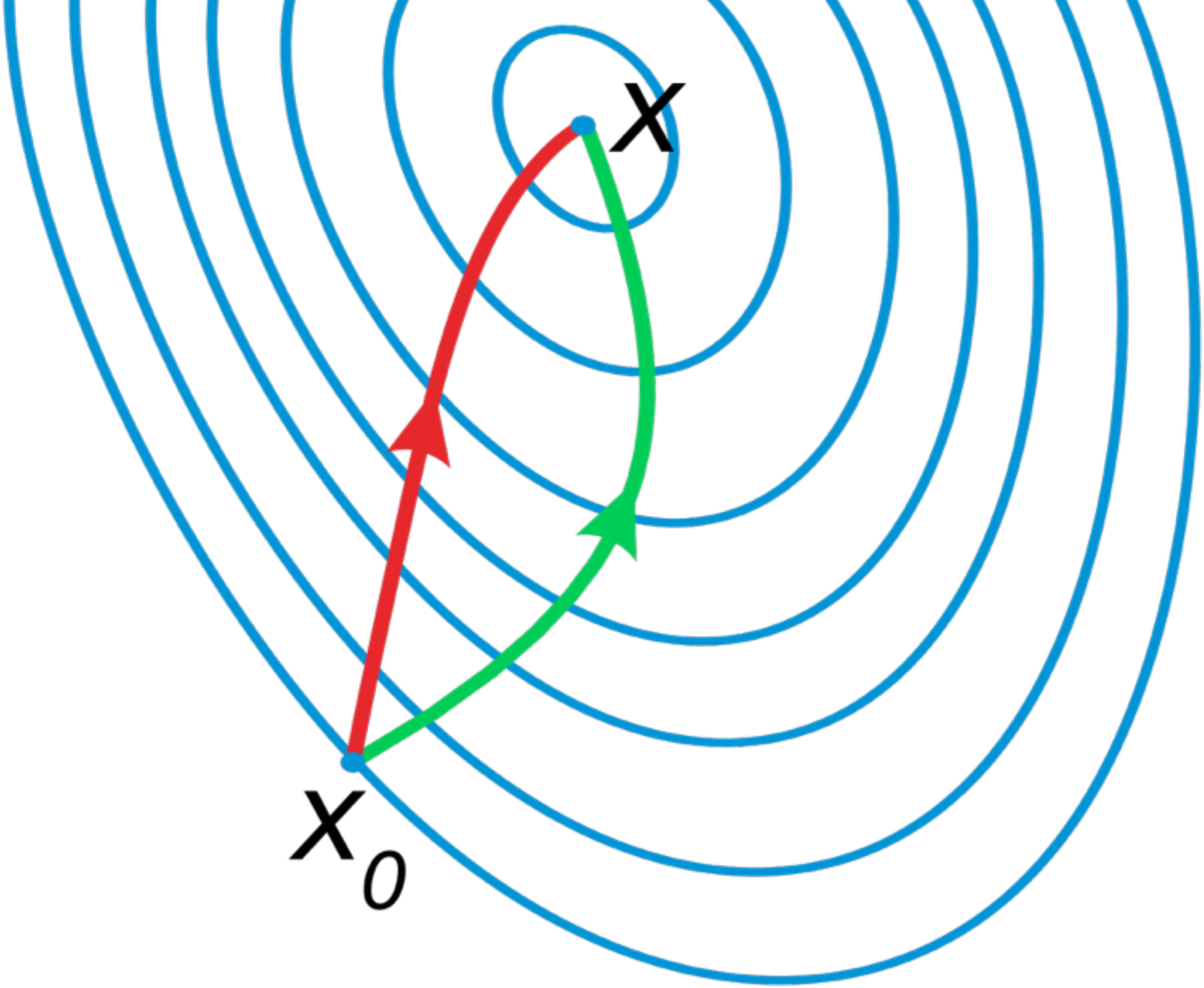


Mathematical optimization

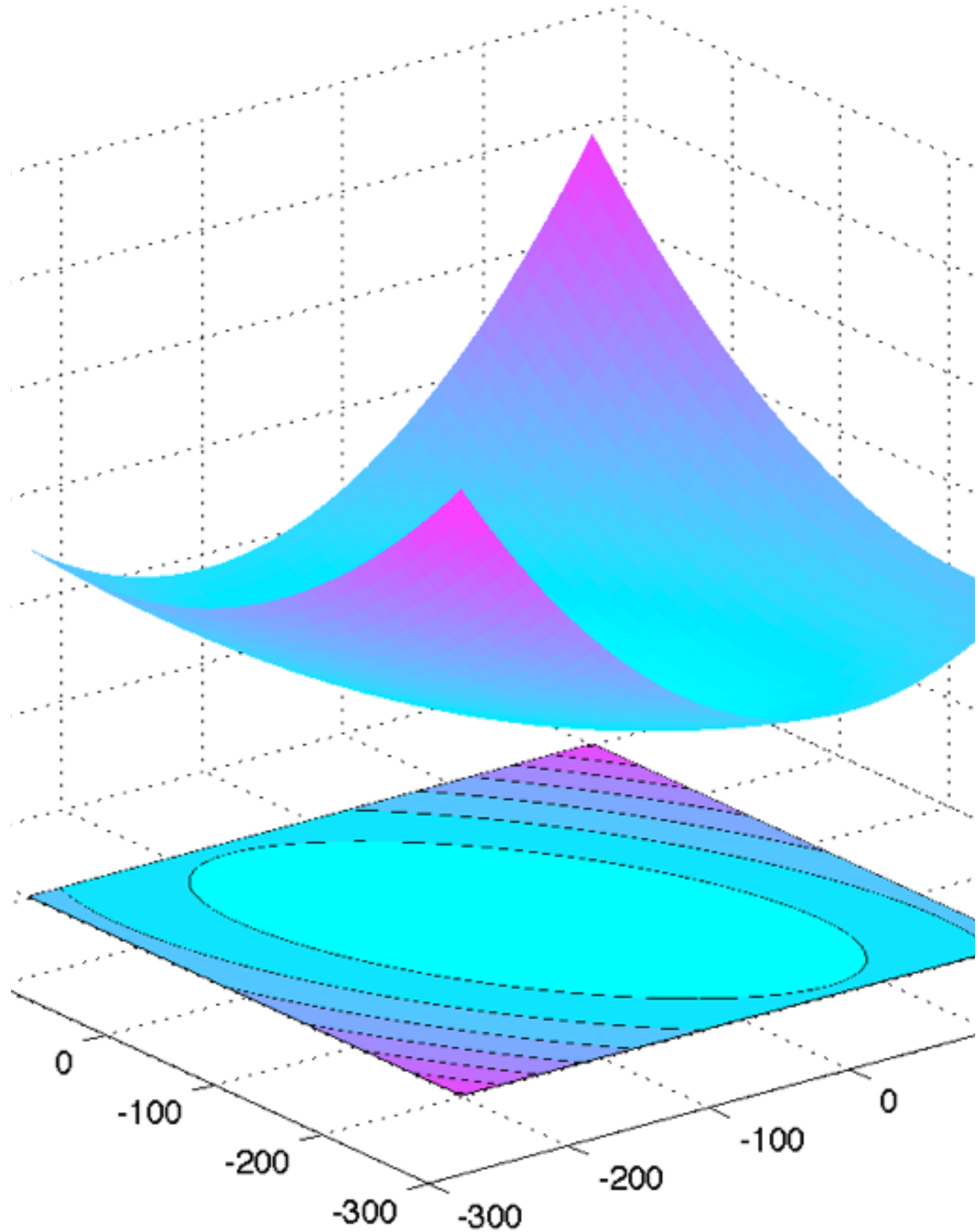
Niru Maheswaranathan
nirum@stanford.edu
4.29.15

What is optimization?

- *Framework:* Conventions for describing or formulating problems mathematically
- *Toolbox:* Algorithms for solving these problems
- *Practical:* Fundamental to many fields: economics, engineering, machine learning, …
- *Mathematics:* Lots of deep and interesting connections between optimization, geometry, statistics

Overview

- Req'd background
    - Calculus (gradients)
    - Linear algebra
- Today:
    - Definitions and examples
    - Convexity
    - Gradient descent
    - Example problem



Definitions

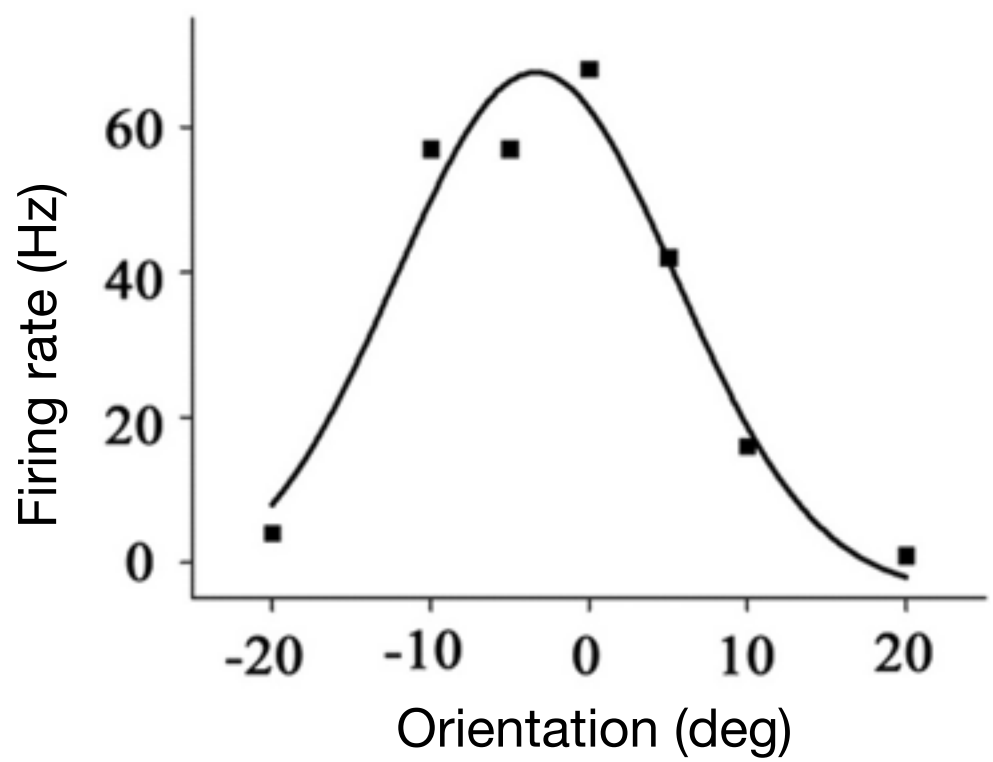
- Objective
- Gradient
- Critical point
- Local optima (minima and maxima)
- Parameter space

Optimization in the life sciences: data fitting

Variables: model parameters

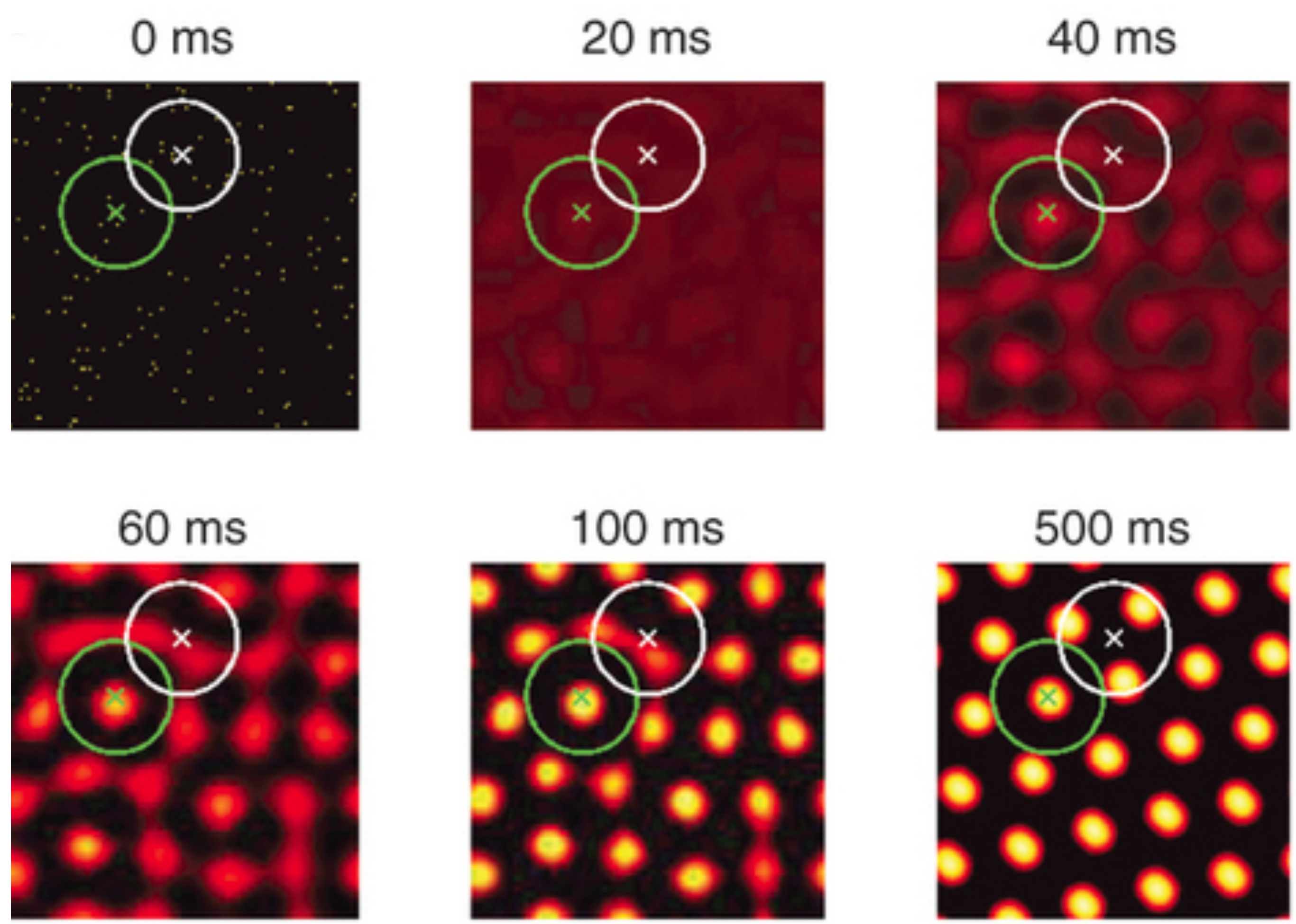
Objective: measure of misfit or prediction error

Fitting a gaussian tuning curve



Examples

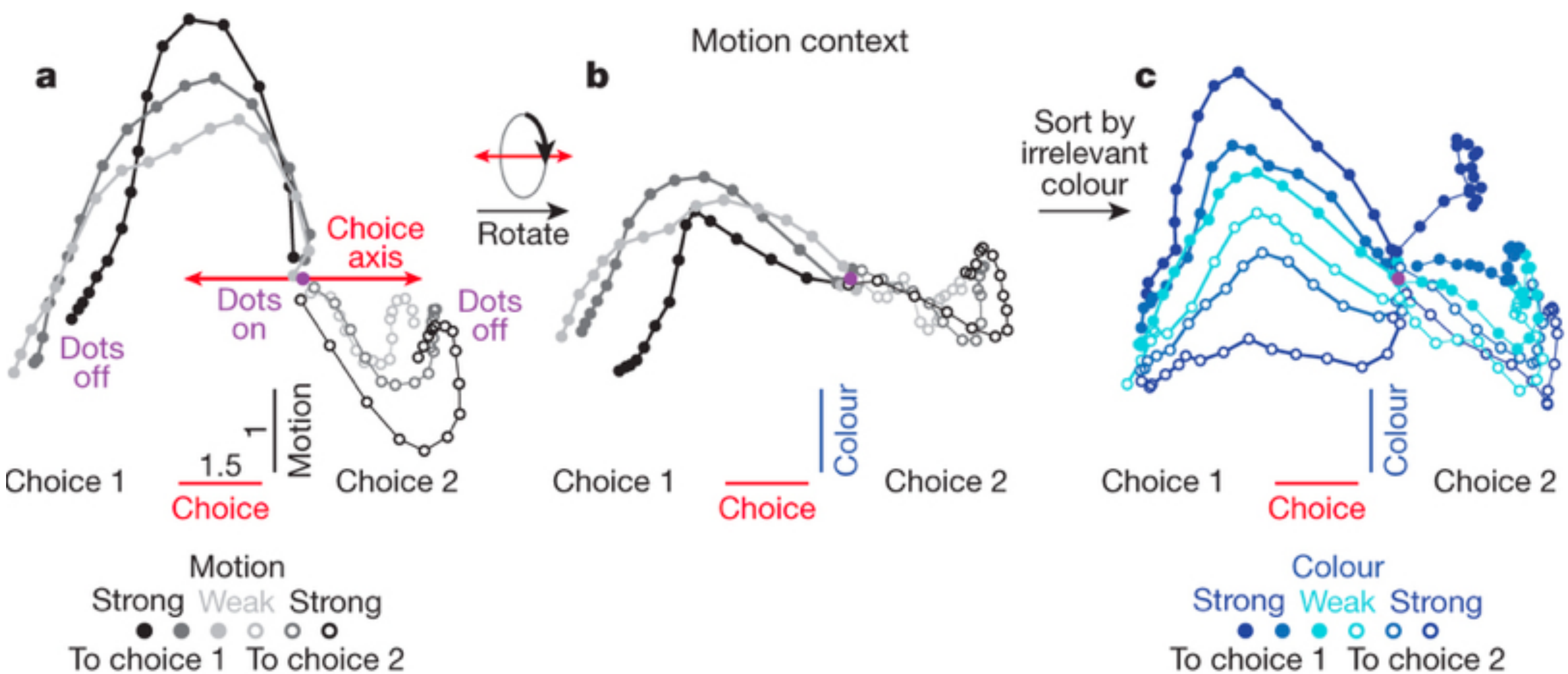
Fitting a grid pattern firing field



Couey et. al. 2013

Examples

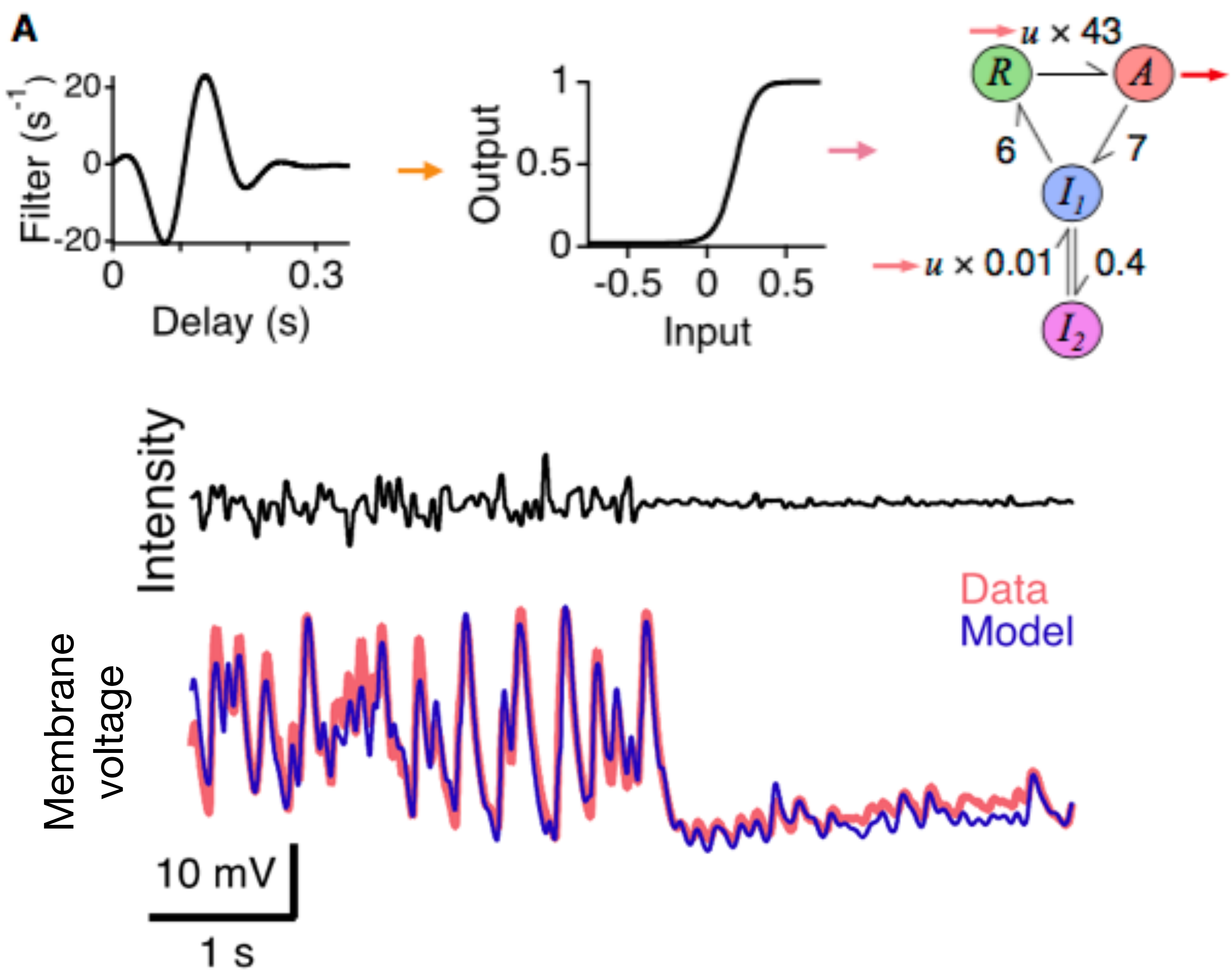
Dimensionality reduction



Mante et. al. 2013

Examples

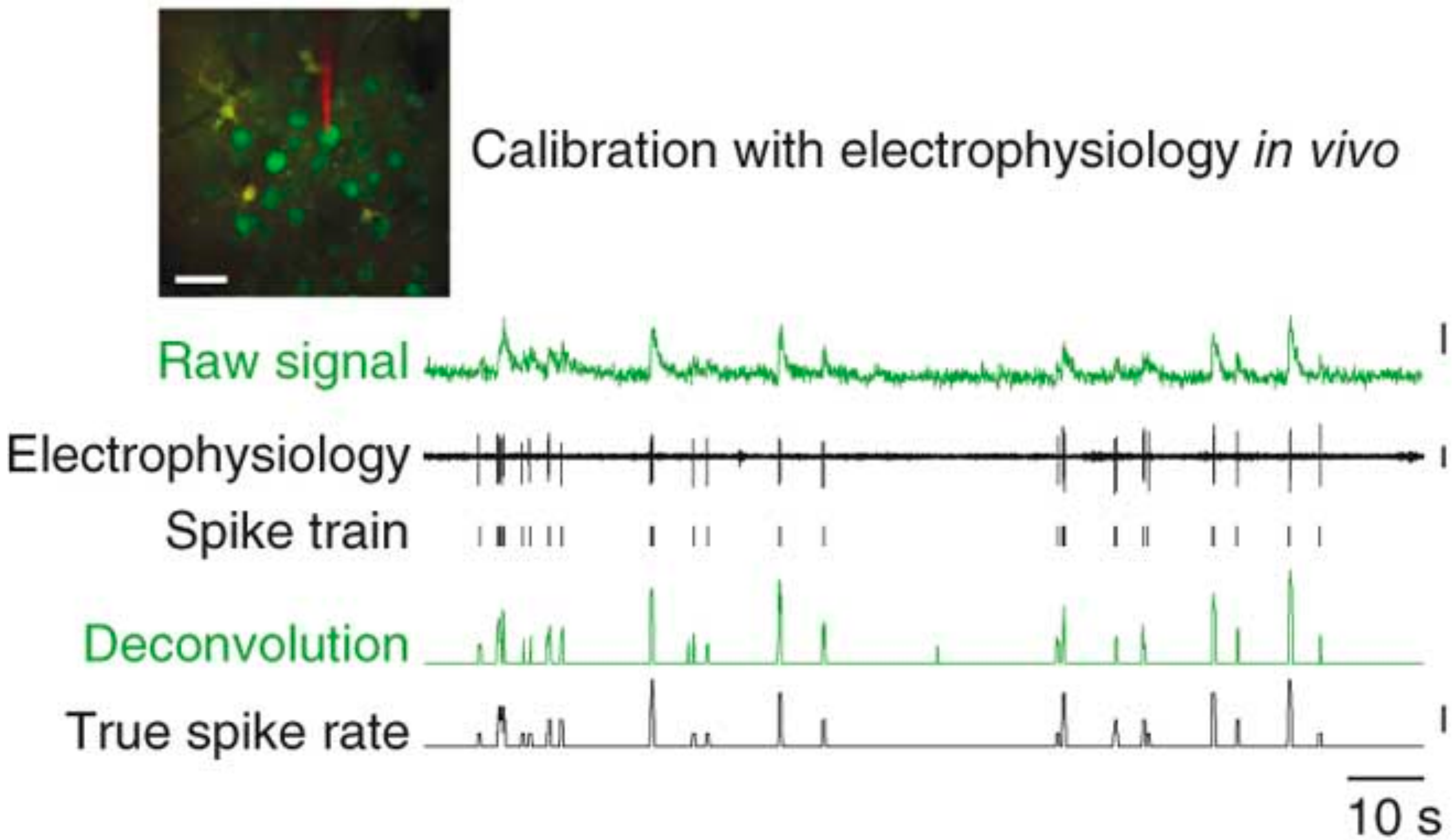
Model fitting



Ozuysal and Baccus 2012

Examples

Imaging data analysis



Smith and Häusser 2010

Exponential improvement in computational power

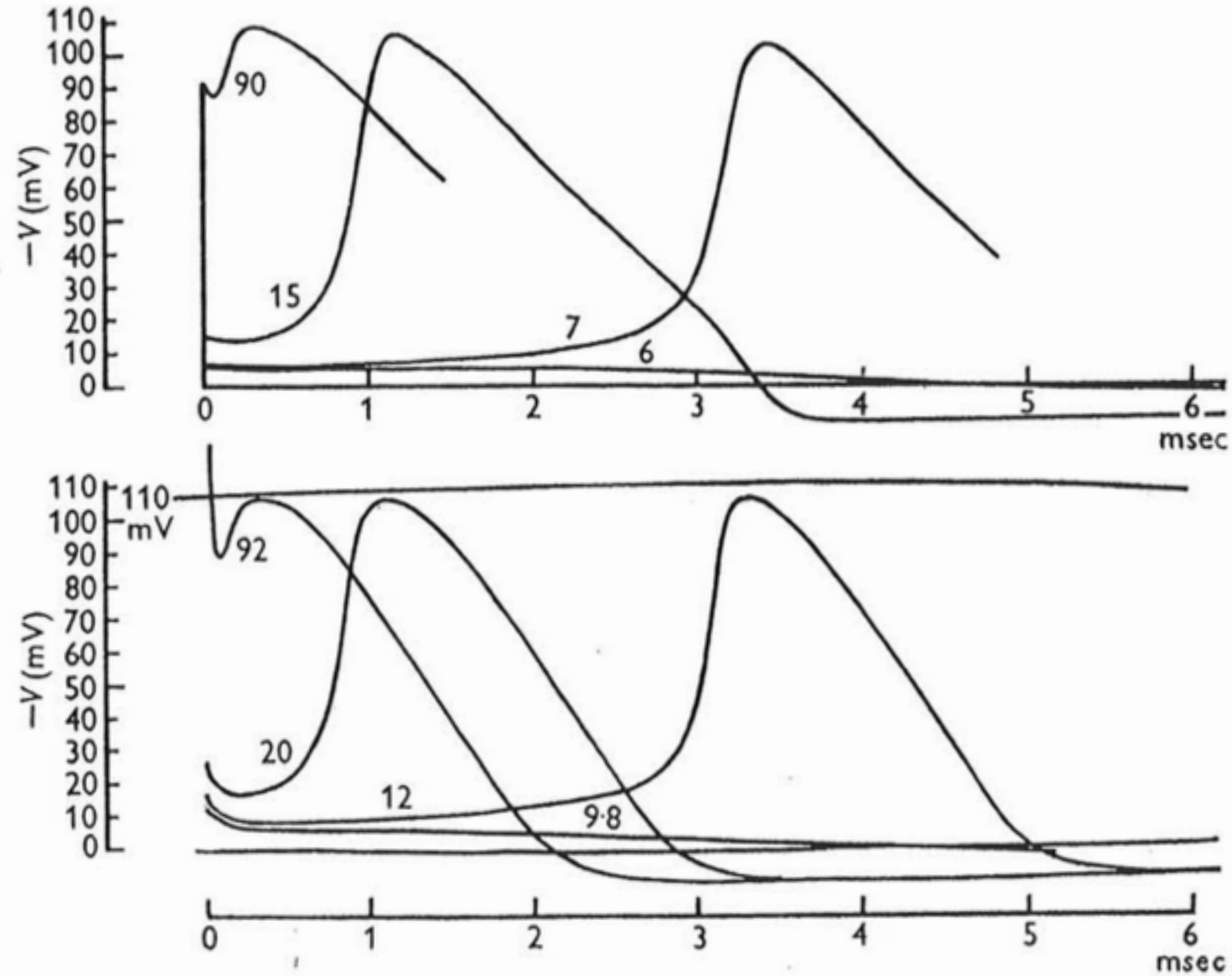


Fig. 12. Upper family: solutions of eqn. (26) for initial depolarizations of 90, 15, 7 and 6 mV (calculated for 6° C). Lower family: tracings of membrane action potentials recorded at 6° C from axon 17. The numbers attached to the curves give the shock strength in $m\mu$coulomb/cm². The vertical and horizontal scales are the same in both families (apart from the slight curvature indicated by the 110 mV calibration line). In this and all subsequent figures depolarizations (or negative displacements of V) are plotted upwards.

Hodgkin & Huxley 1952

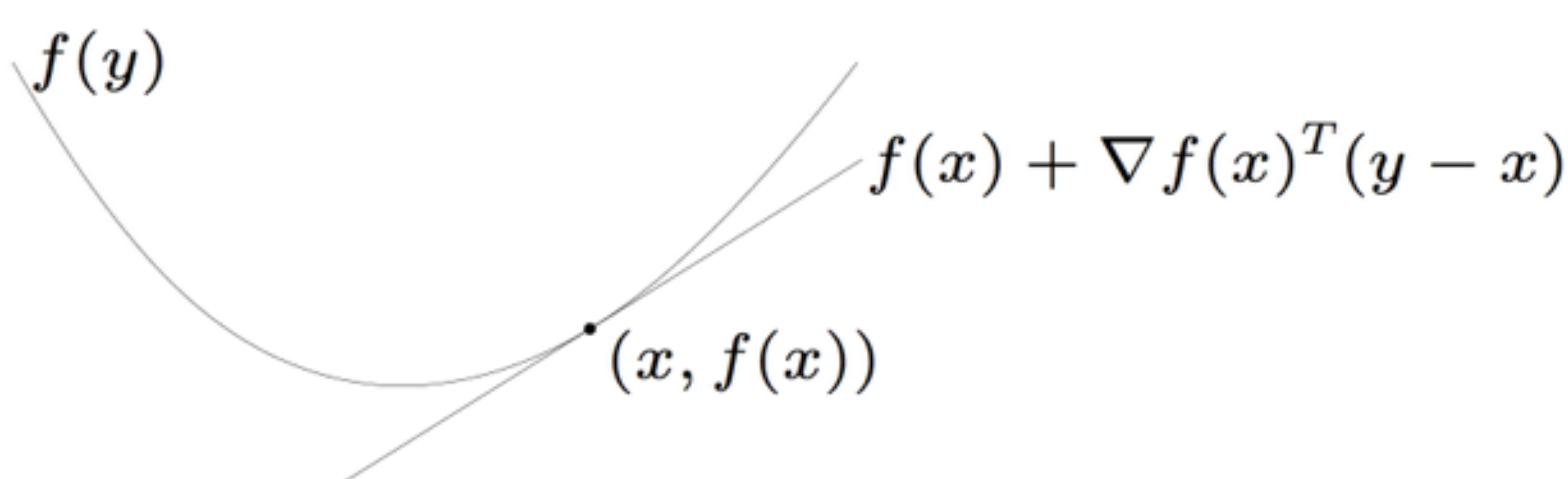
Convexity

A function *f* is convex if:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$

Convexity

- Guarantees a global optimum for the objective
- Turns local information into global information
- Efficient algorithms for convex problems, can solve them exactly

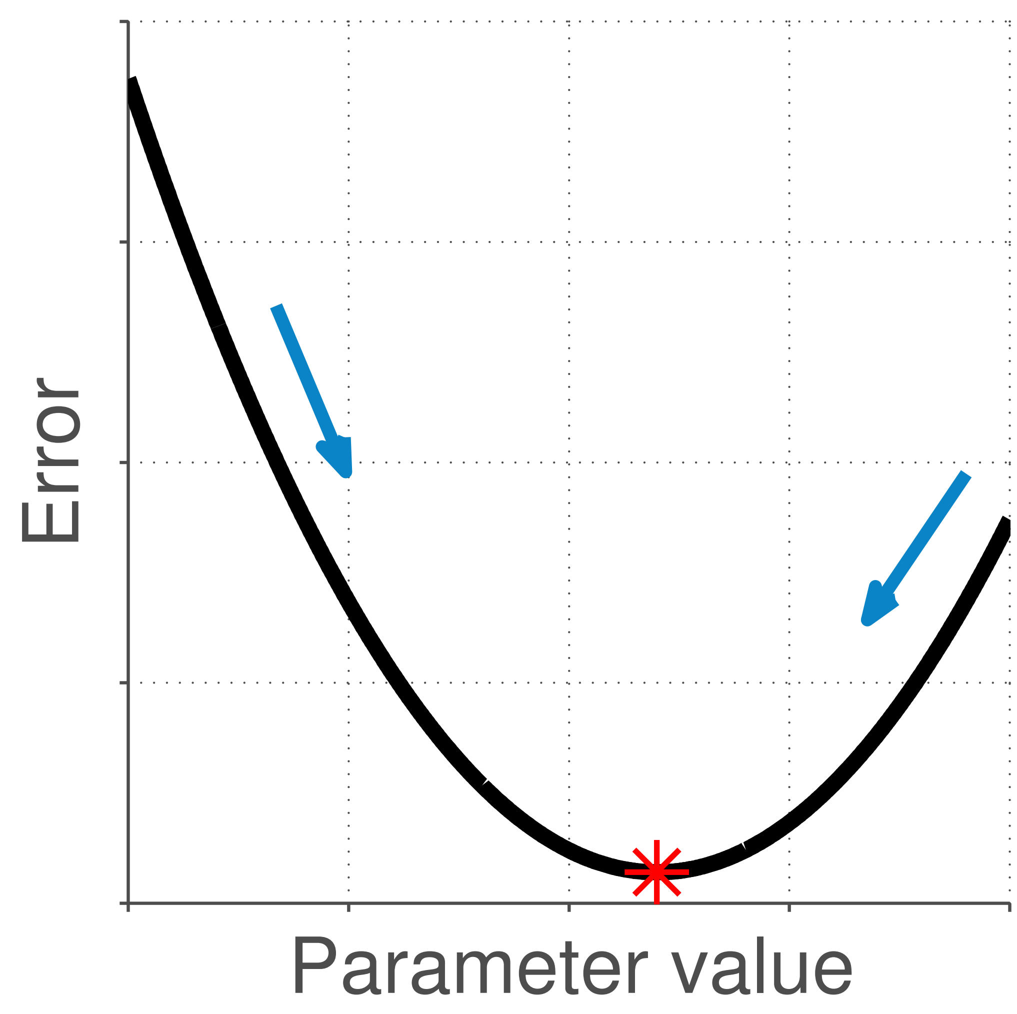


first-order approximation of f is global underestimator

Gradient descent

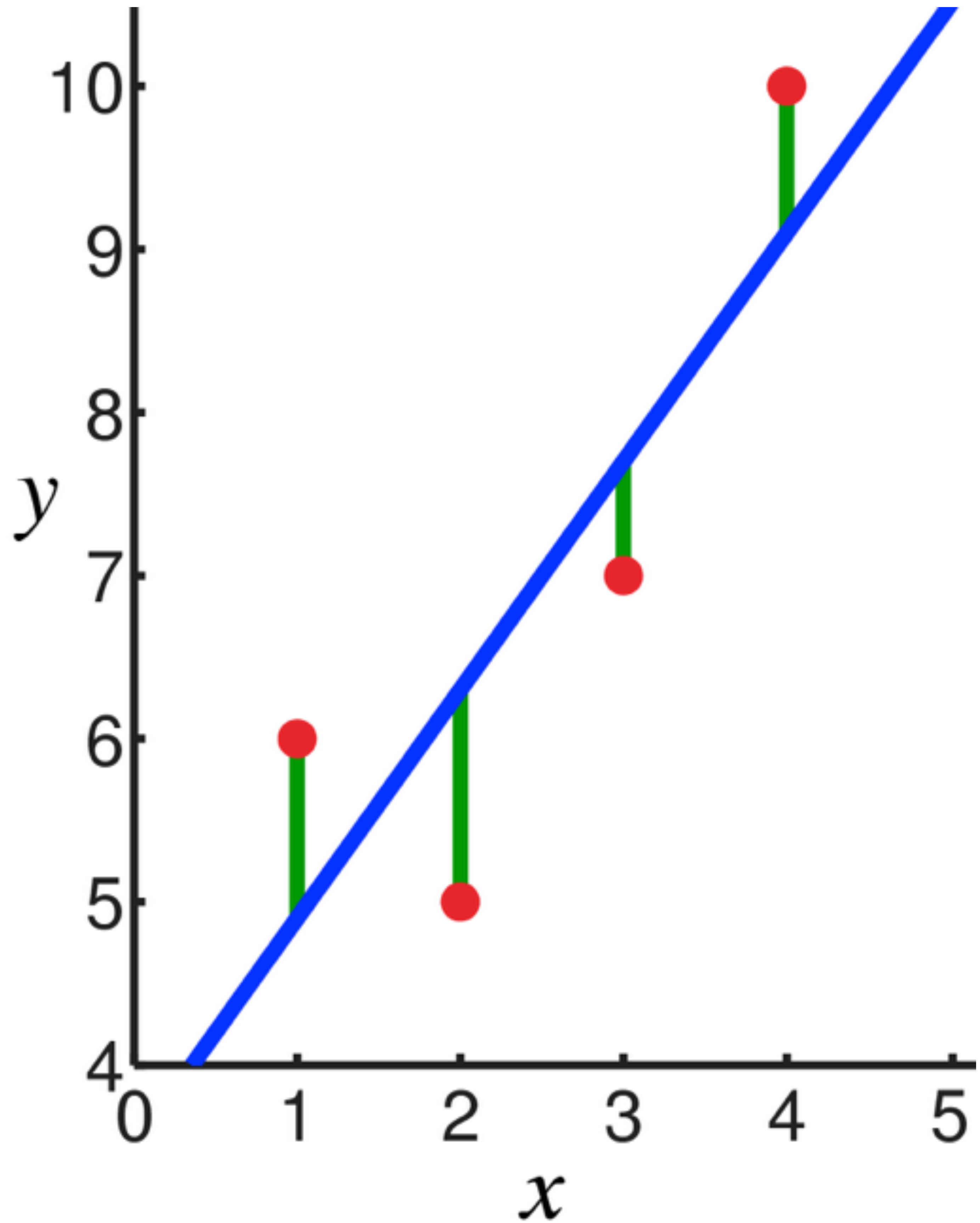
Gradient descent: choose initial $x^{(0)} \in \mathbb{R}^n$, repeat:

$$x^{(k)} = x^{(k-1)} - t_k \cdot \nabla f(x^{(k-1)}), \quad k = 1, 2, 3, \ldots$$

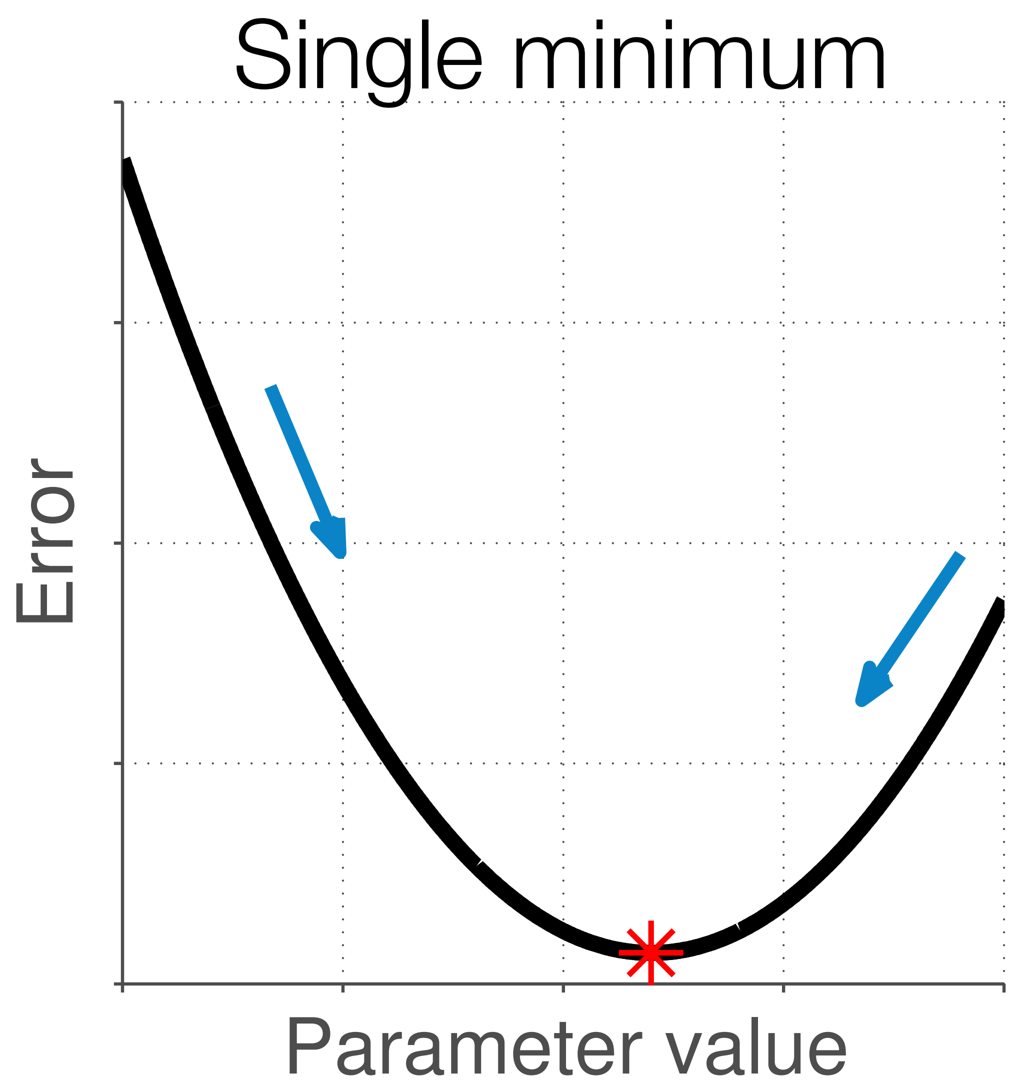


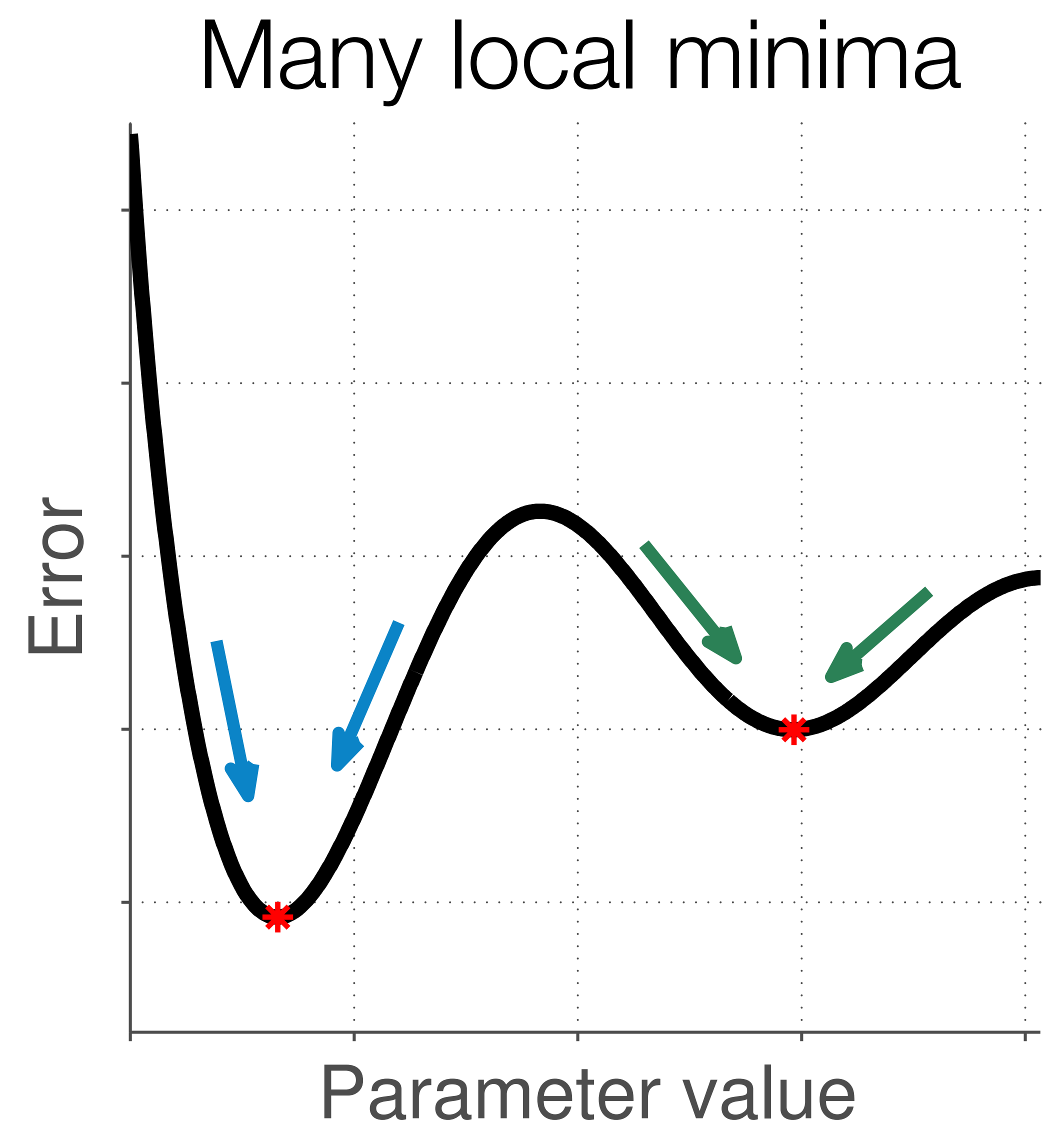
Least squares

- Data: pairs of points (x,y)
- Objective: Minimize the squared error between the observed data and a linear function
- Variables: parameters (mean and slope) of the line
- Demo



Problems with local minima





Take home messages

- Optimization: recognize when it comes up
- Definitions: objective, gradient, optima, parameter space
- Convexity: definition, why it's useful
- Gradient descent: why it works, why it ought to work