Principles of Parameter Estimation

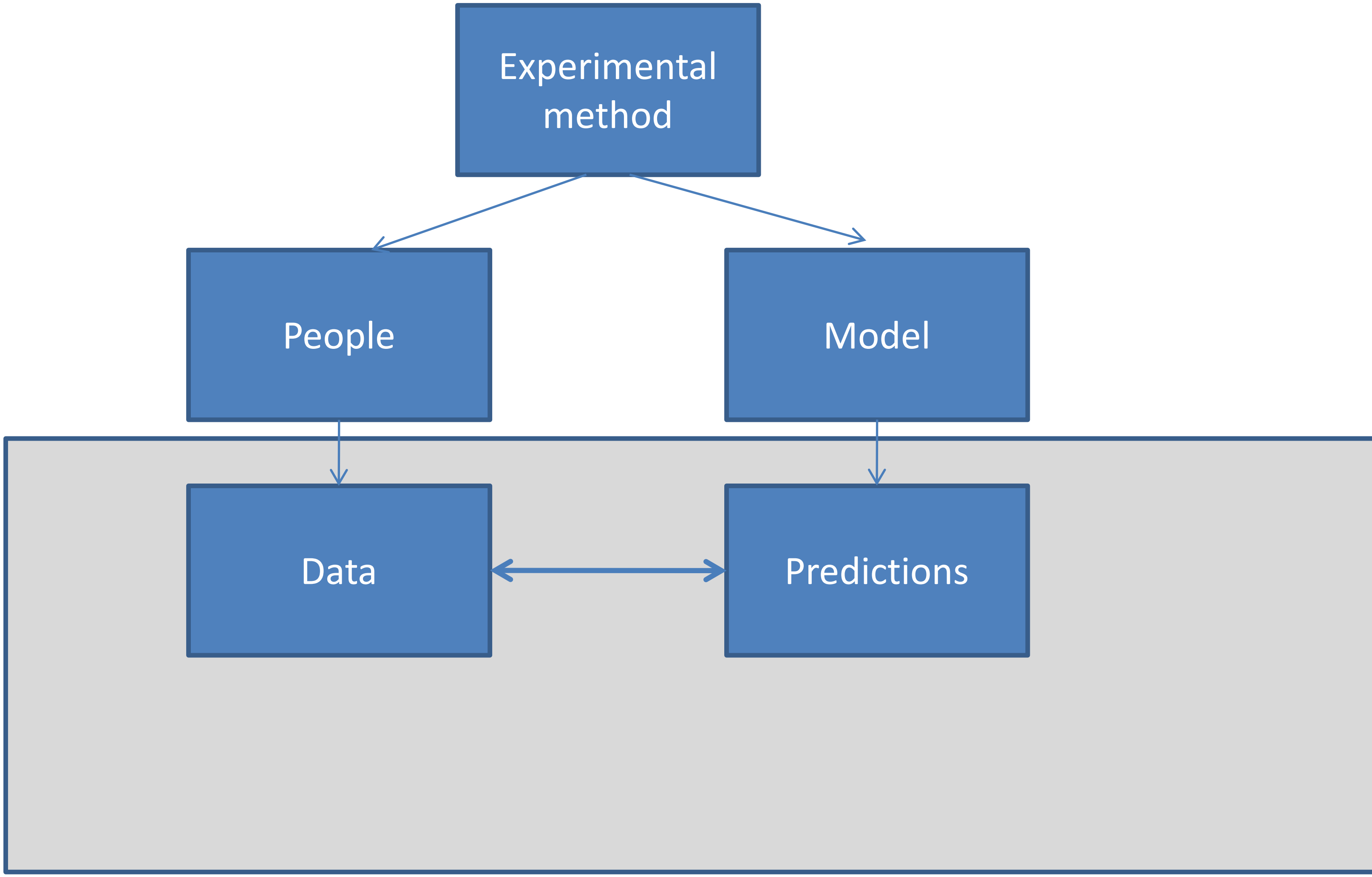
Stephan Lewandowsky

Overview for Today (Wednesday)

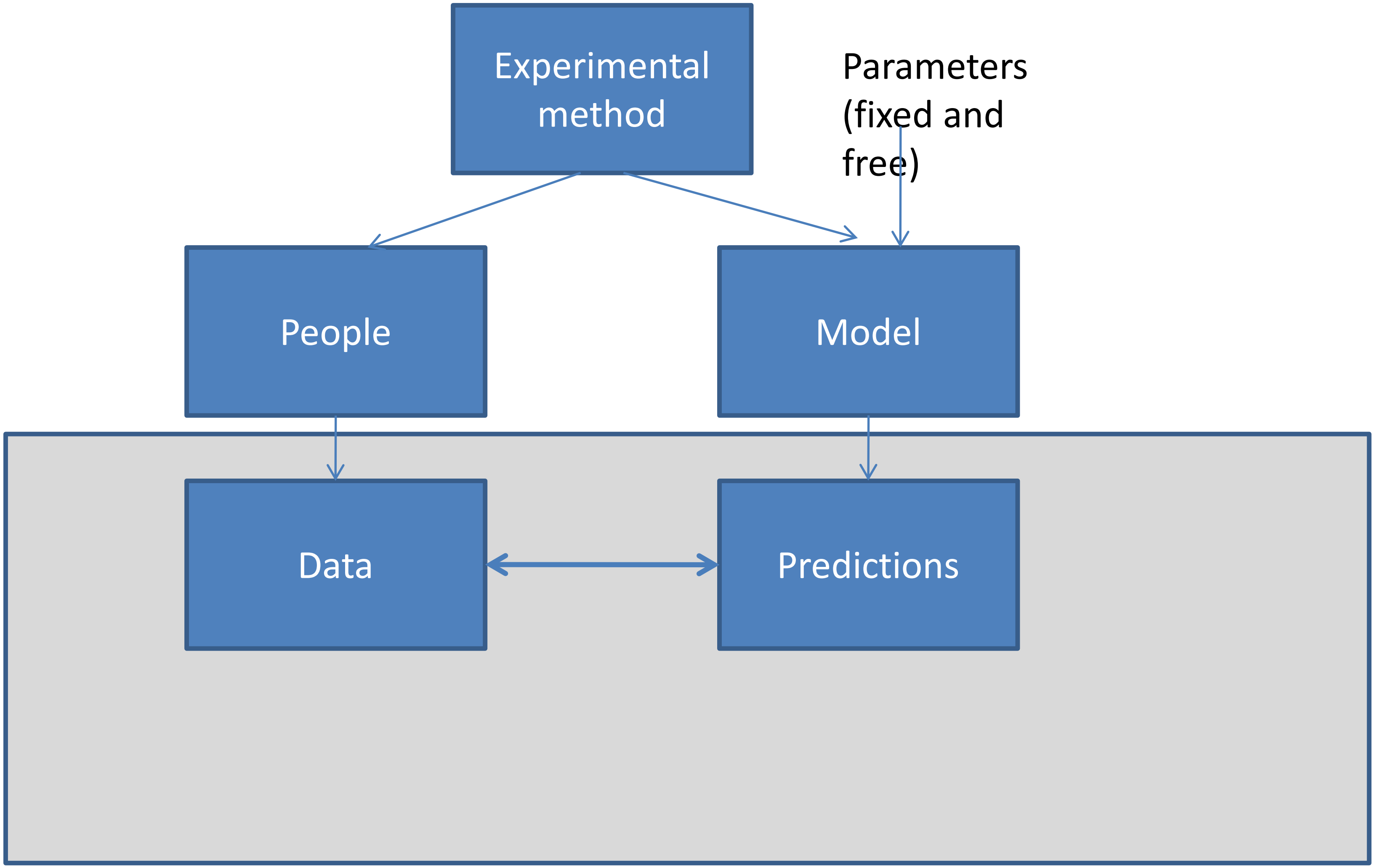
Time	Instructor	Topic
9:00 – 10:30	Stephan Lewandowsky	Parameter estimation
10:30 – 1:00	Gordon Brown, Cas Ludwig, Laura F.	From basics to maximum likelihood estimation
1:00 – 2:00		lunch
2:00 – 4:00	GB, CL, LF	From basics to maximum likelihood estimation, exercises
4:00 – dinner	All	meet mentors and discuss projects
6:30/7:00		Dinner
8:00	All	Bowling

Yes, there will be breaks

Objectives of Modeling



Objectives of Modeling



Classes of Parameters

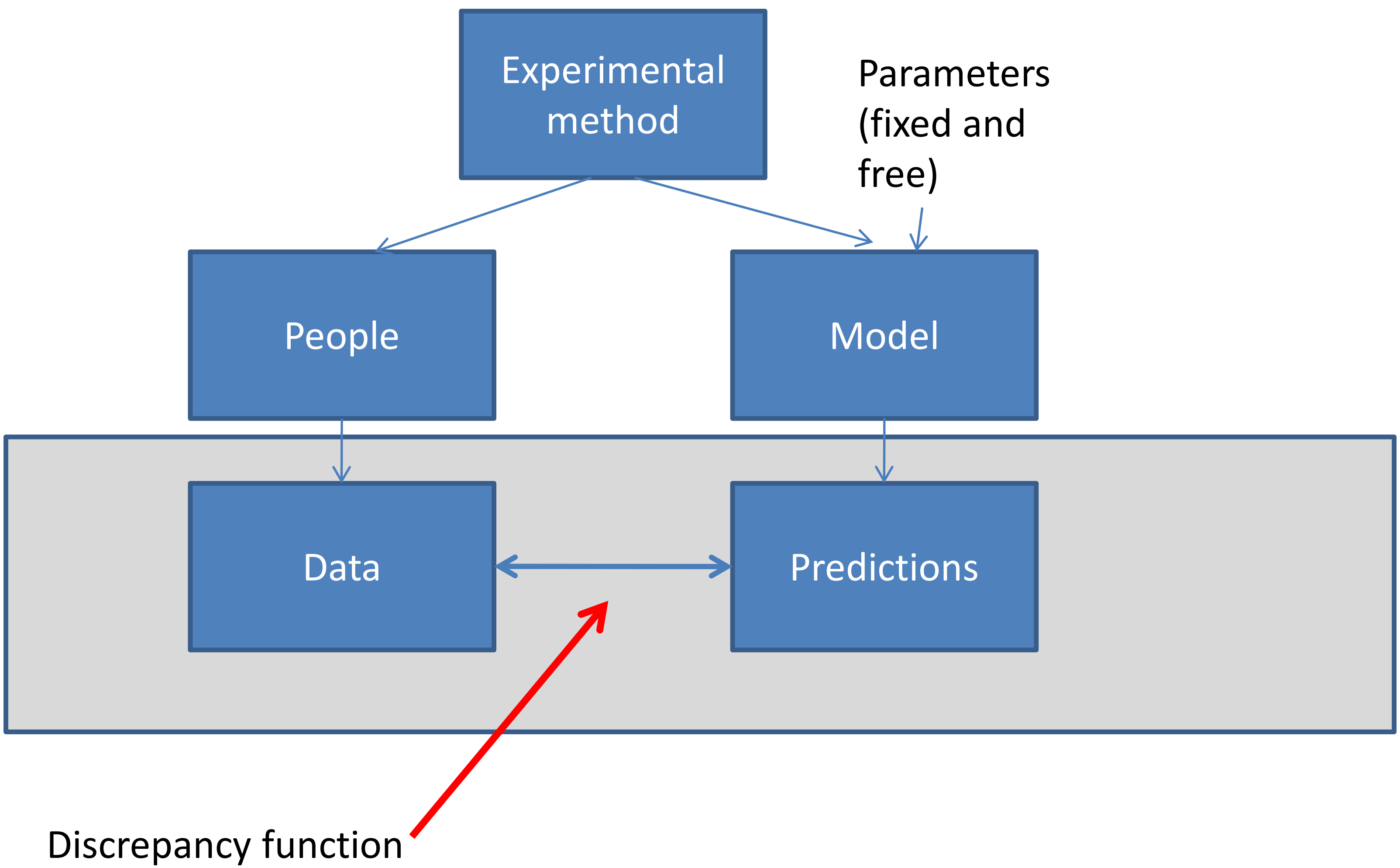
- Free parameters
  - estimated from the data to maximize fit of model
  - parsimony of model determined (in part) by their number
  - the more free parameters, the greater the flexibility (all other things being equal)
- Fixed parameters
  - not estimated but fixed
  - less relevant to determine parsimony

Free or Fixed?

```
#random walk model
nreps <- 1000
nsamples <- 2000

drift <- 0.0  #noninformative stimulus
sdrw <- 0.3
criterion <- 3
```

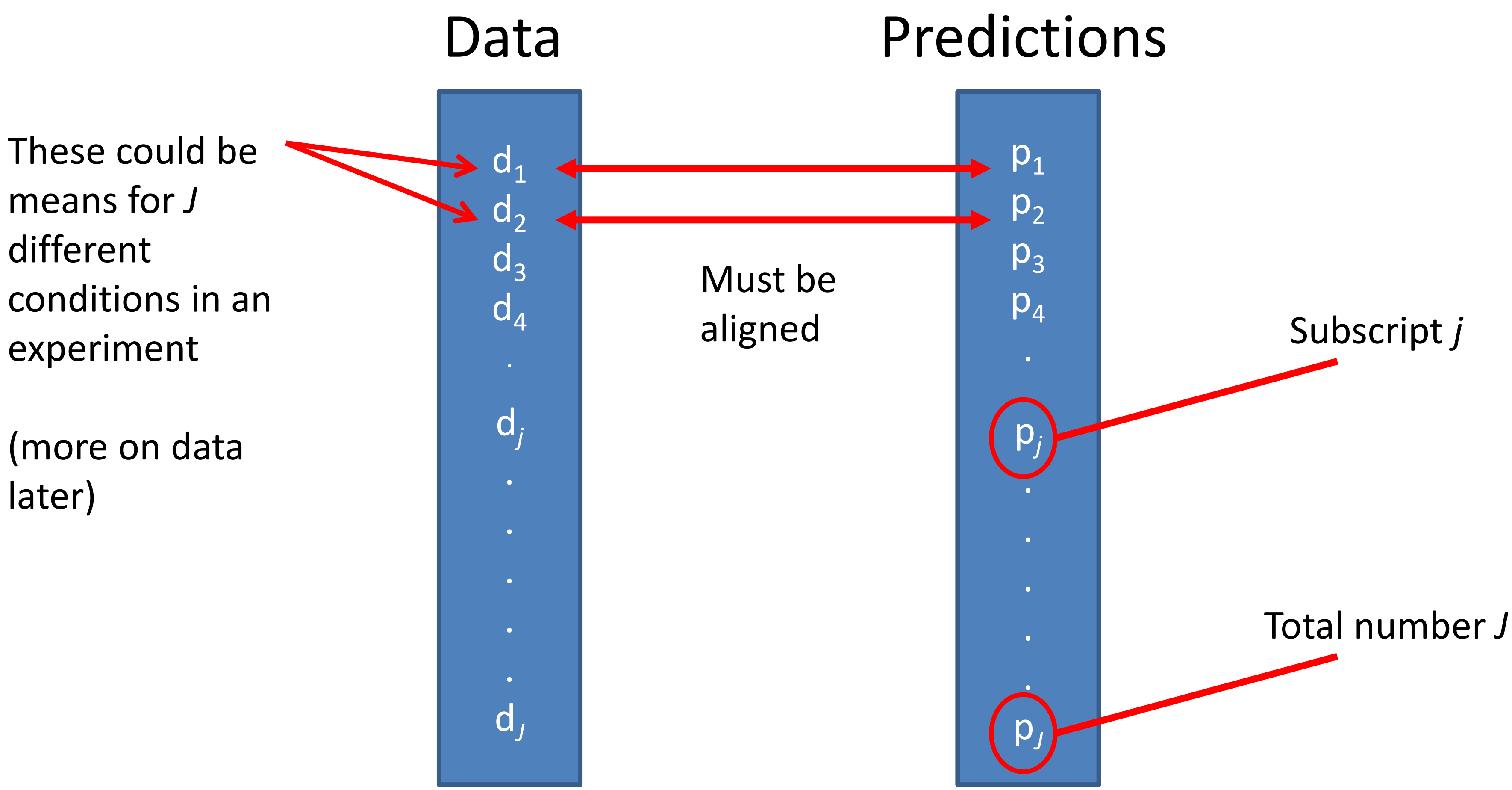
Discrepancy Function



Discrepancy Function

- Requirements
  - must yield single numeric value
  - must be continuous
- By convention
  - express model fit as discrepancy ...
  - ... that is to be minimized
- Other labels
  - cost function, error function, objective function

Discrepancy Function

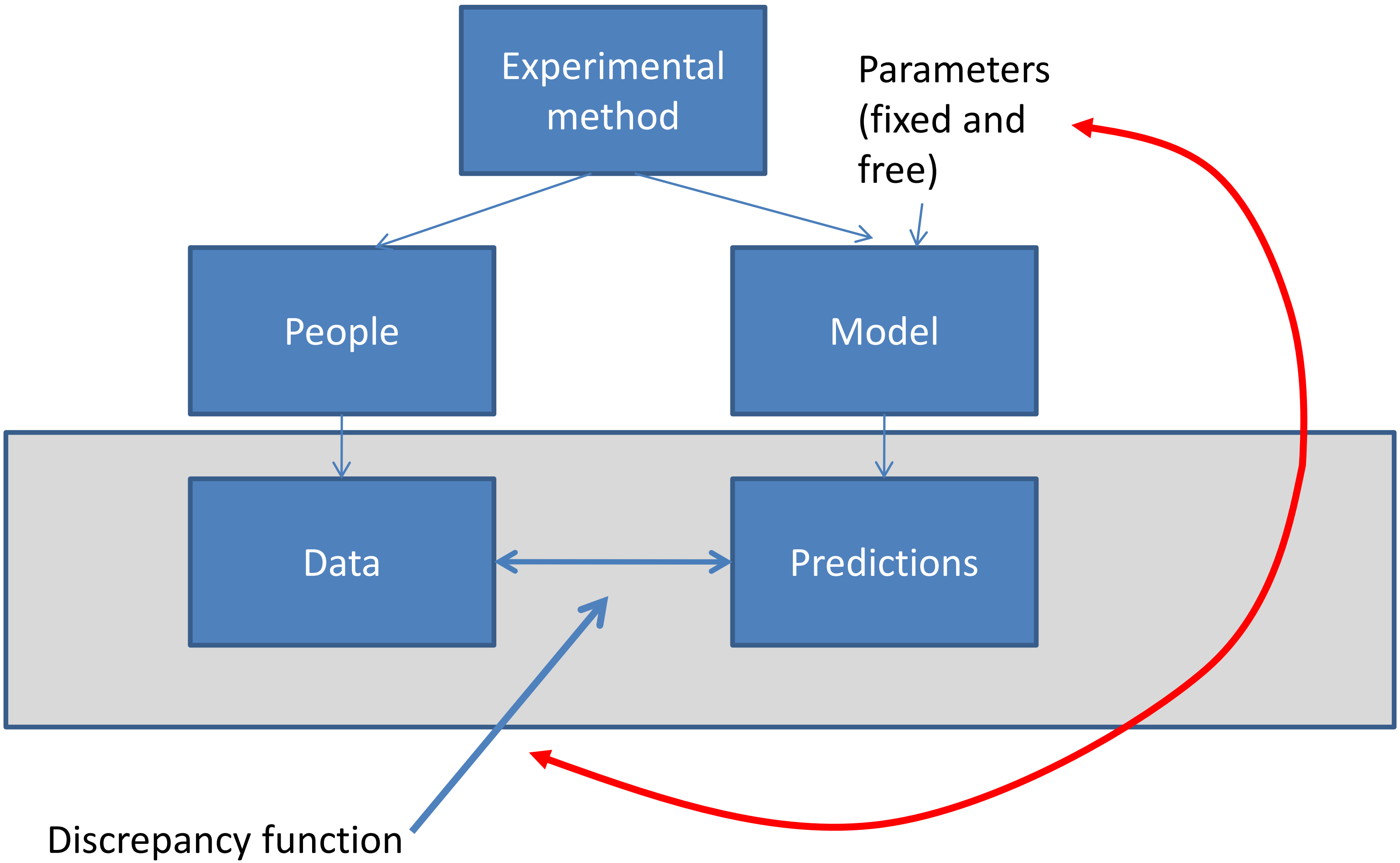


RMSD

- Residual|Root Mean Squared Deviation ("least squares")

$$RMSD = \sqrt{\frac{\sum_{j=1}^{J} (d_j - p_j)^2}{J}}$$

Parameter Estimation

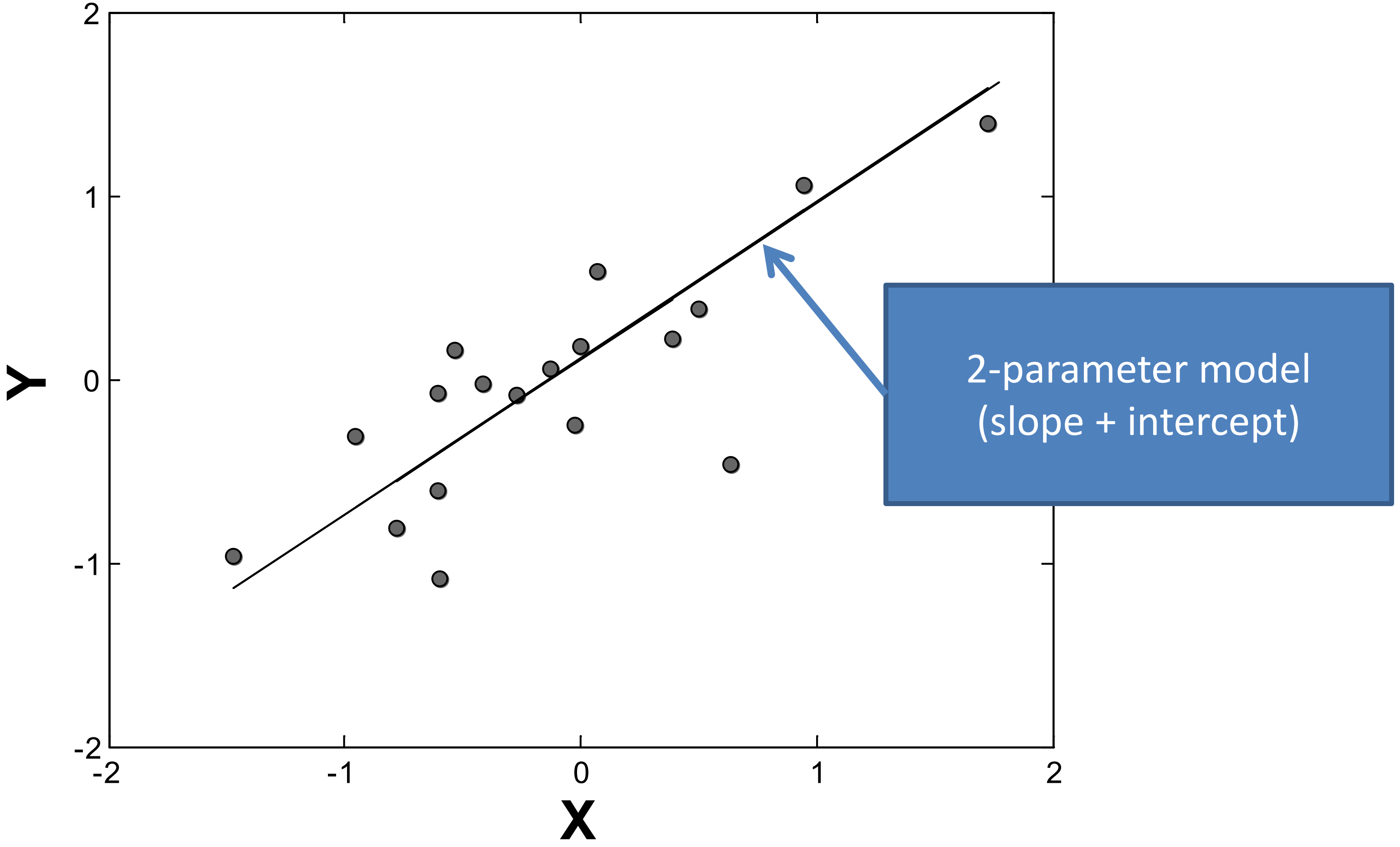


Parameter Estimation

- The choice of discrepancy function is independent of the choice of parameter-estimation technique
- RMSD
  - usually has no obvious statistical properties
  - parameter estimates have no obvious statistical properties
  - ... unlike maximum likelihood (ML: later today)
- But: everything that follows also applies to ML

Introductory Example

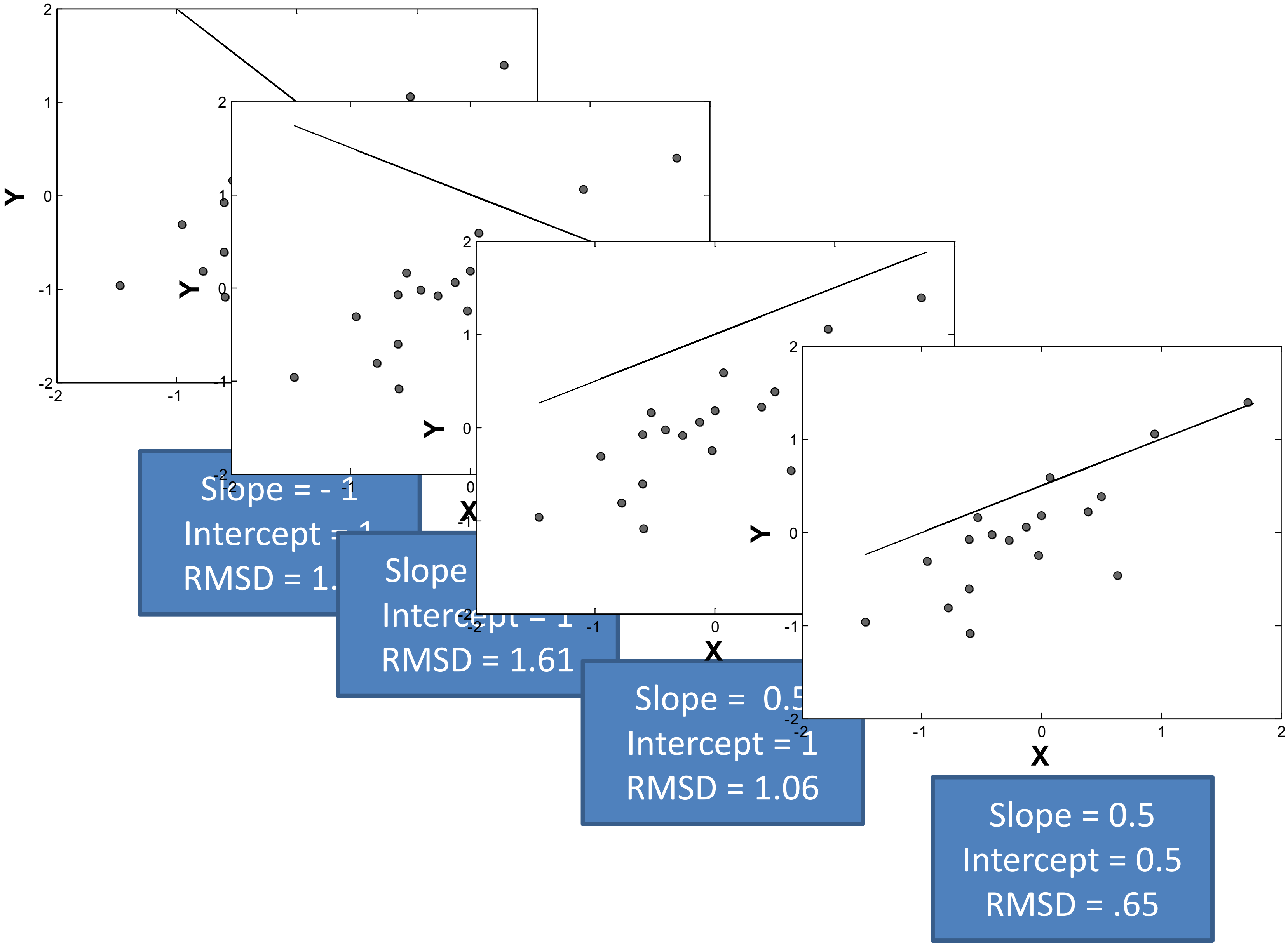
Introductory Example



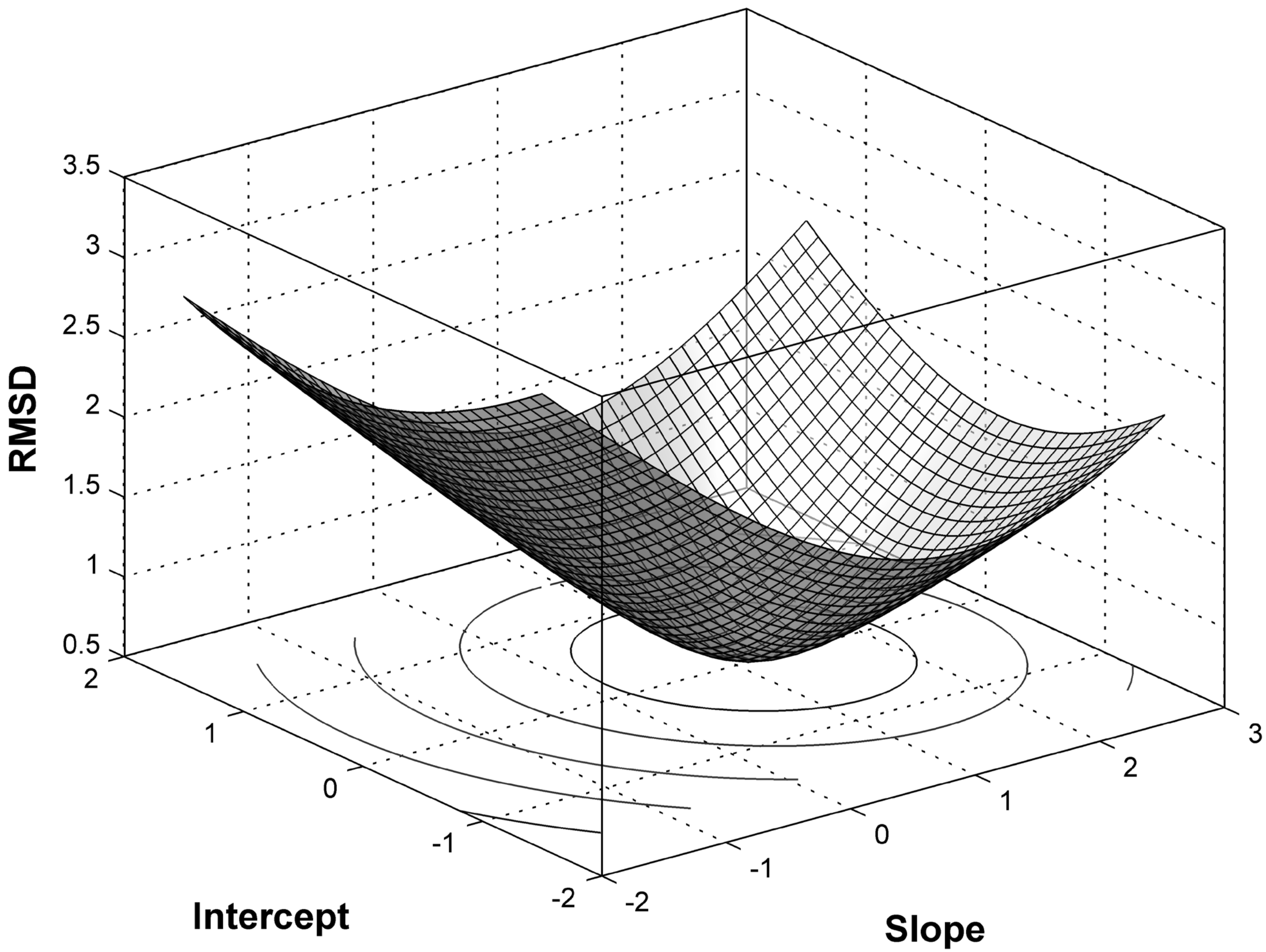
Simple Linear Regression

- Why this particular regression line?
  - computed by analysis program
  - “least squares” solution

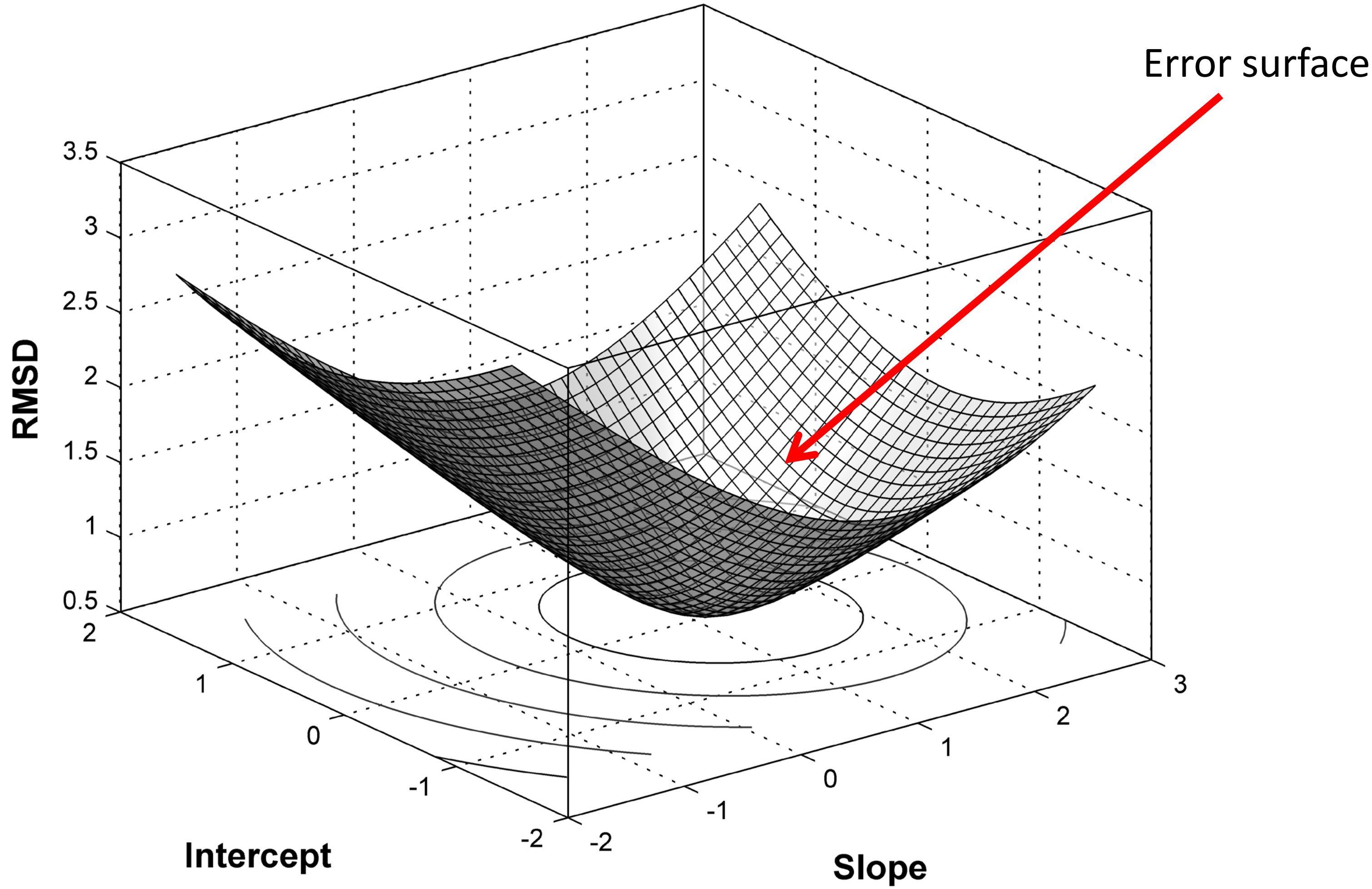
- Let’s explore some others


Slope = - 1
Intercept
RMSD = 1.
Slope
RMSD = 1.61
Slope = 0.5
Intercept = 1
RMSD = 1.06
Slope = 0.5
Intercept = 0.5
RMSD = .65
X
Y

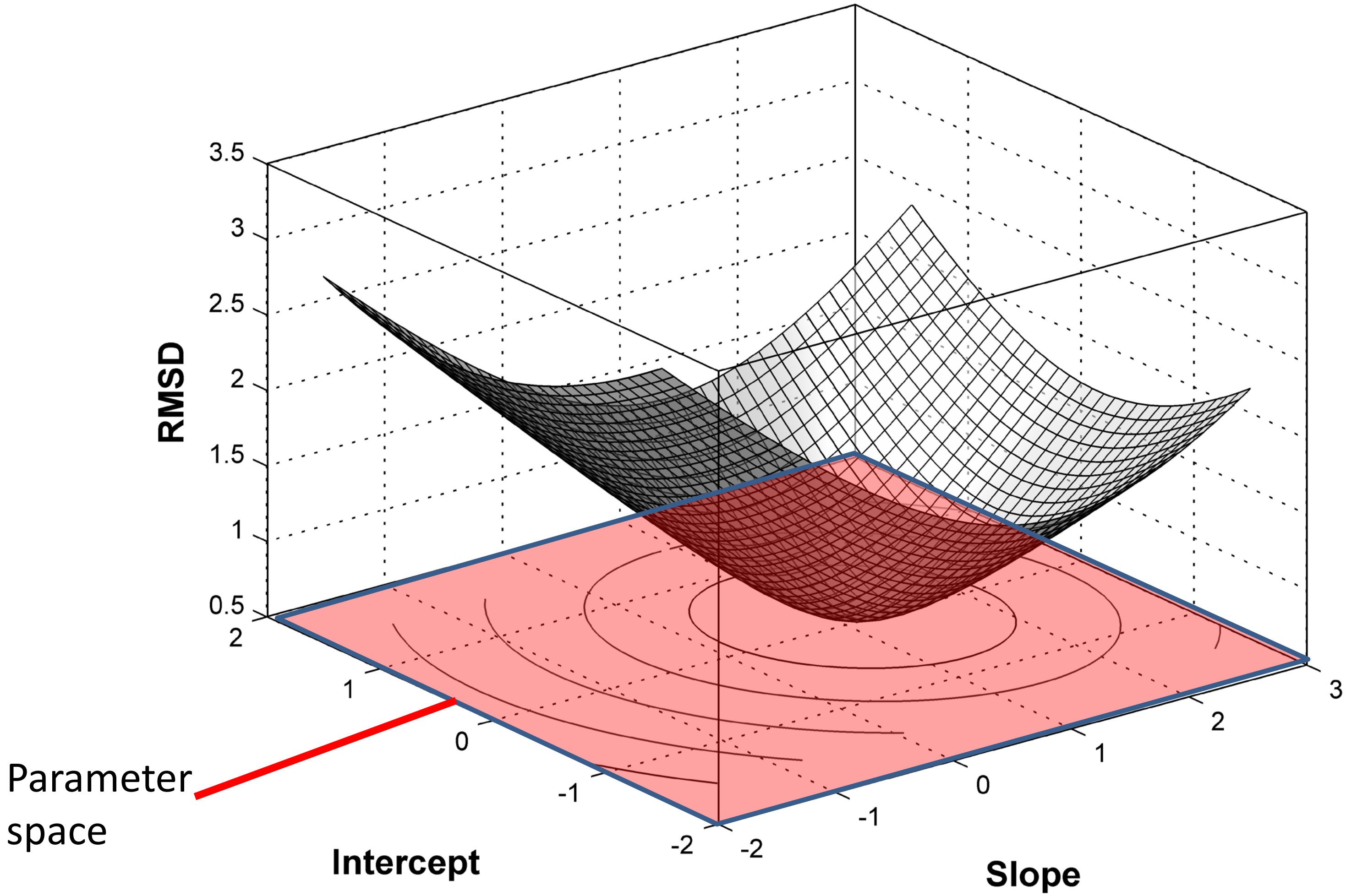
Considering All Parameter Values



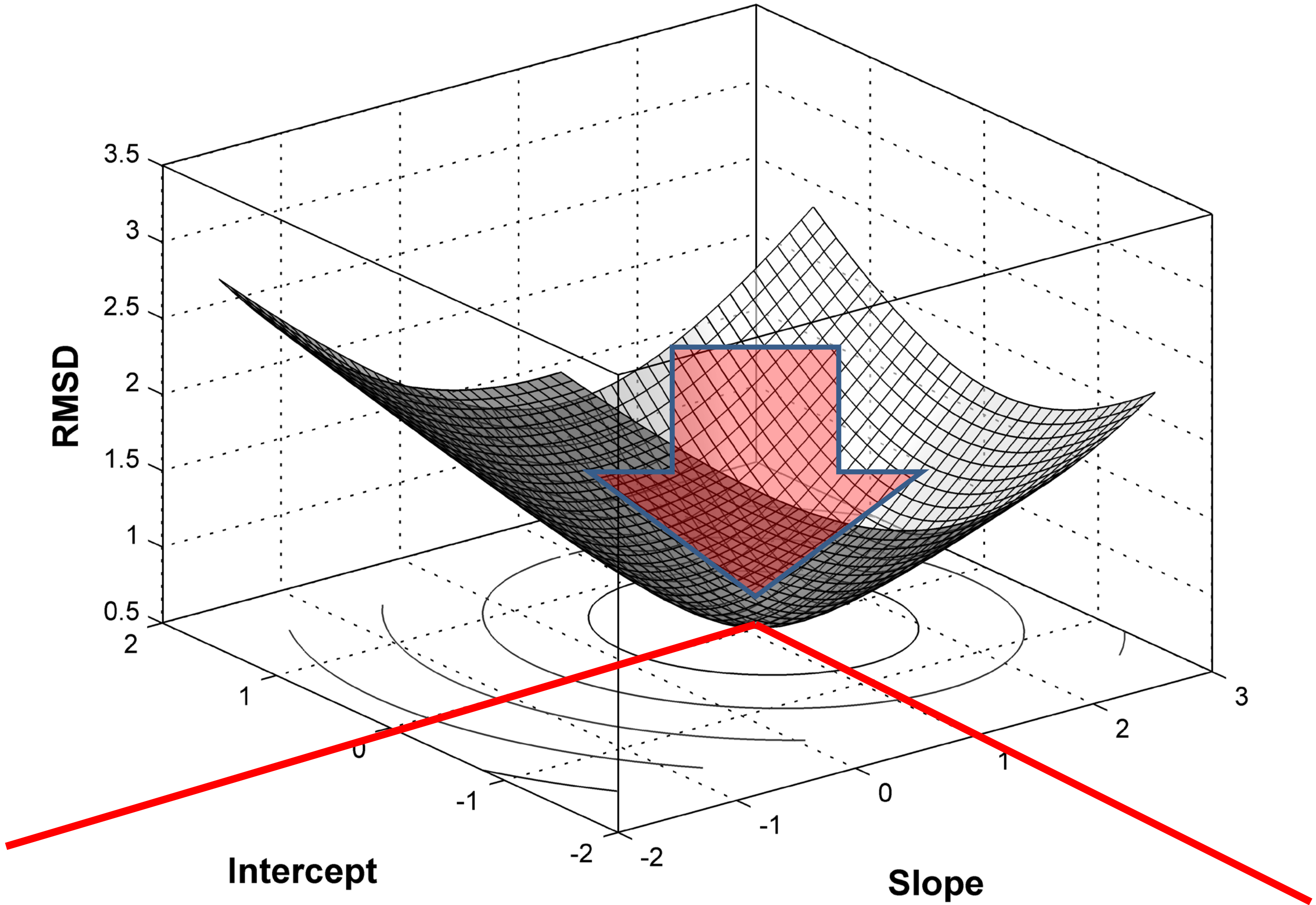
Considering All Parameter Values



Considering All Parameter Values



Considering All Parameter Values



Parameter Estimation

- Move through parameter space ...
- ... from some set of starting values ...
- ... down the error surface ...
- ... until a minimum is reached.
- Those are the best-fitting parameter estimates.

How?

Parameter Estimation Techniques

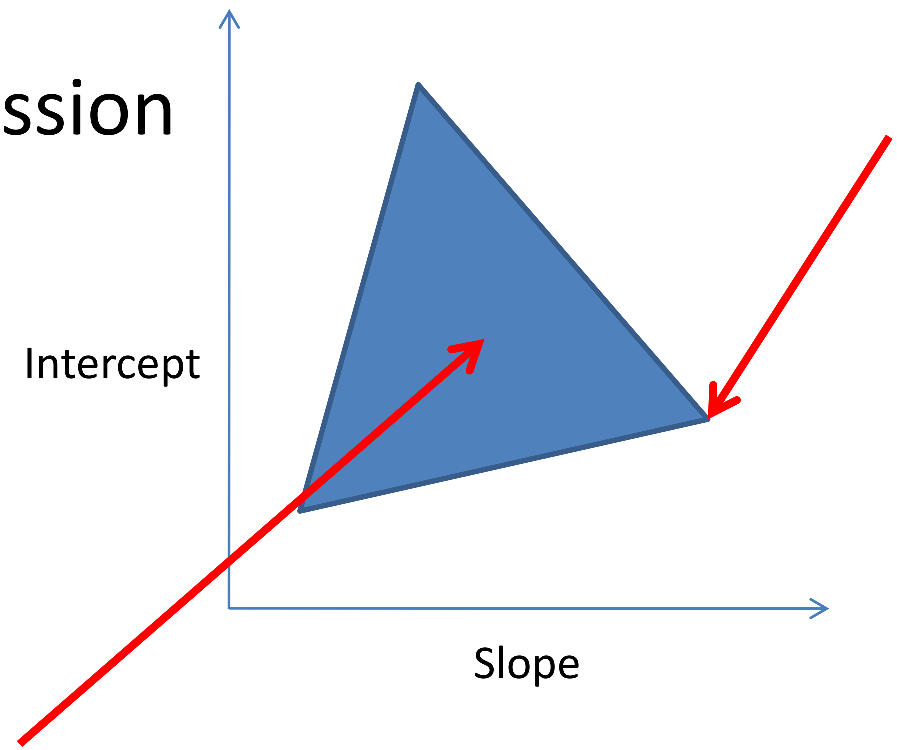
- Many methods exist
- We focus on SIMPLEX
  - others are in the text book
- Assume that parameter values are continuous

SIMPLEX

- Simplex = set of D+1 interconnected points for arbitrary dimensionality D
  - 2D = triangle
  - 3D = pyramid
  - 4D = pentachoron
- In SIMPLEX, dimensionality = number of parameters
  - 2 parameter regression = triangular simplex

SIMPLEX

- Regression

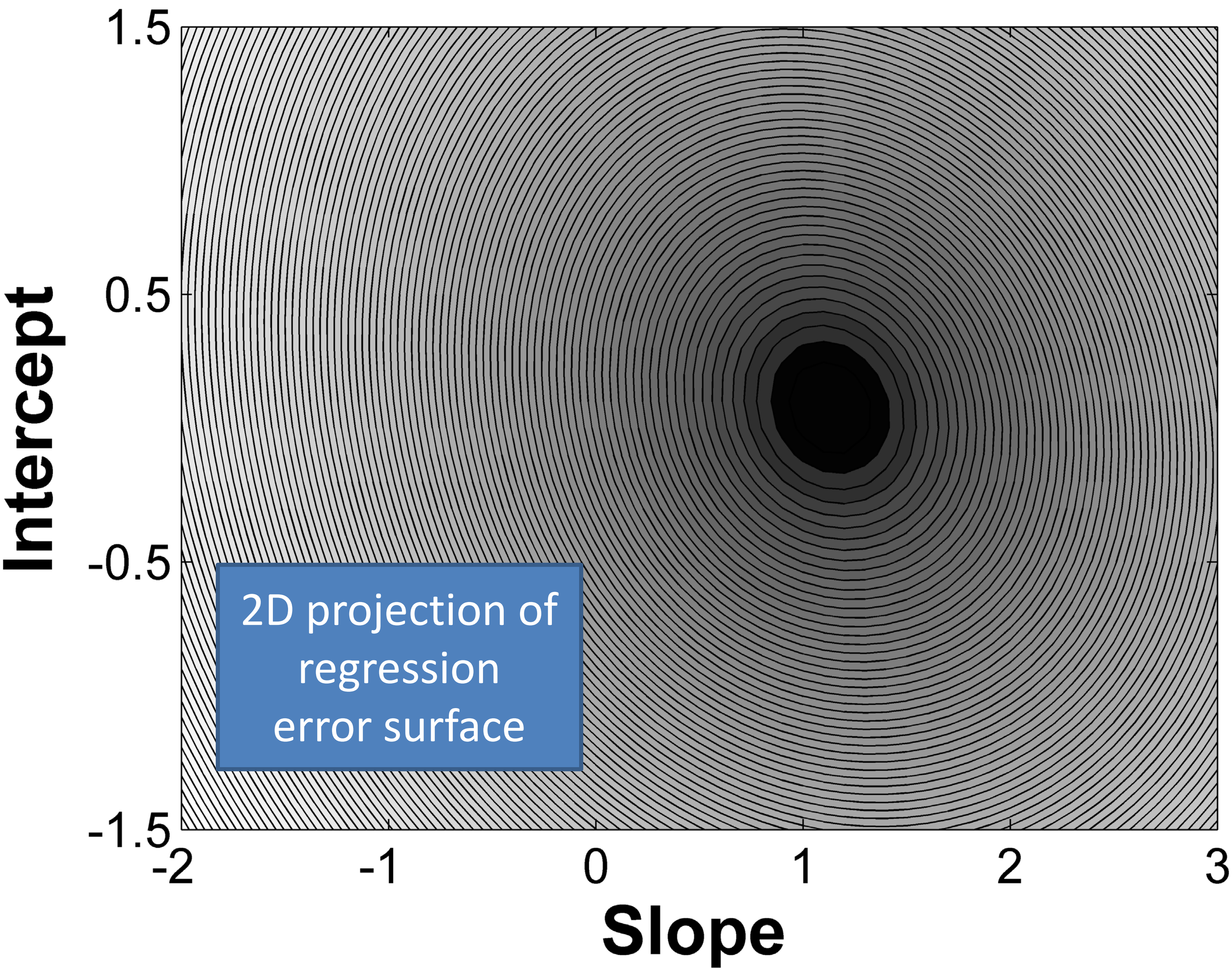


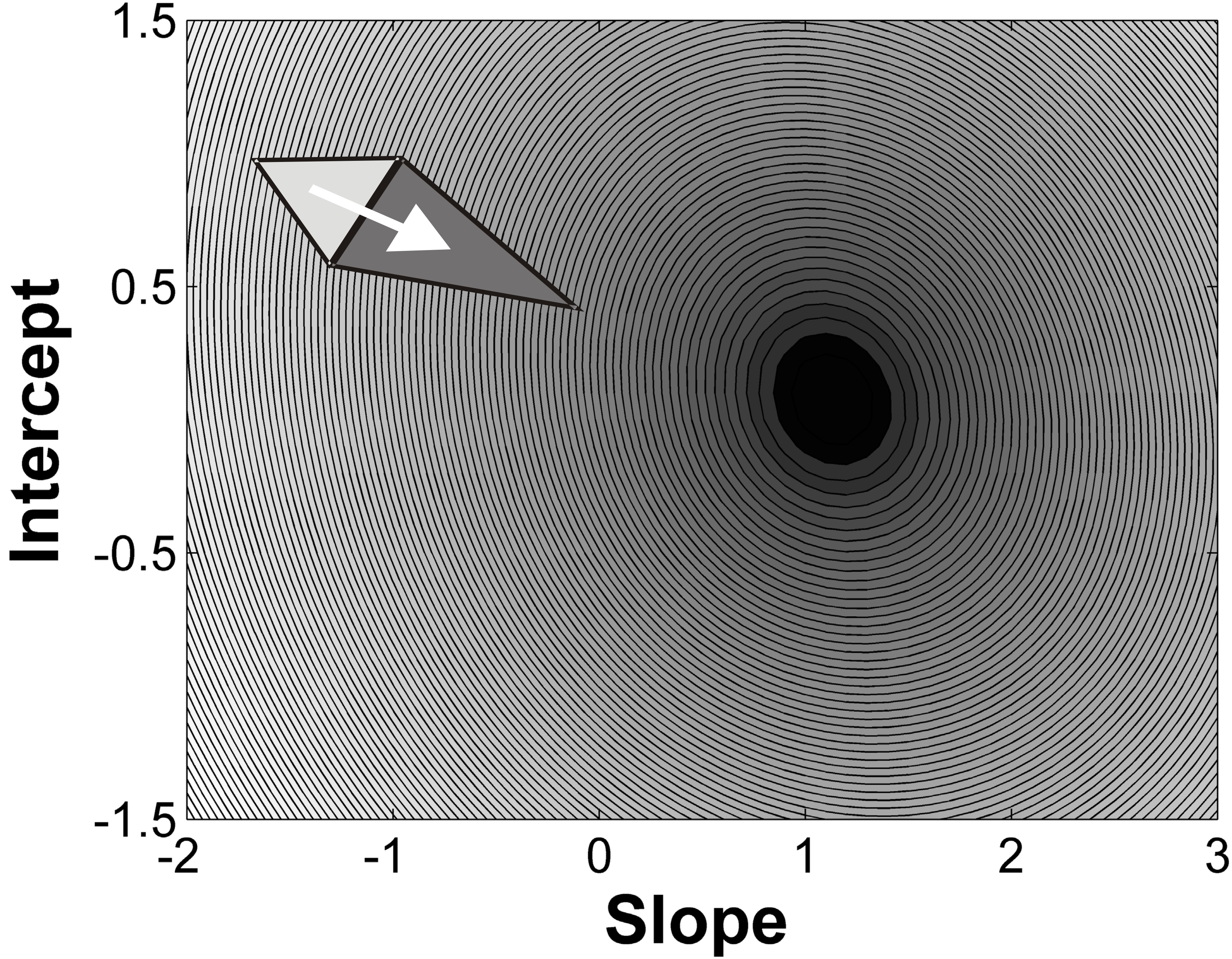
each vertex corresponds
to set of parameter values

created around the starting values

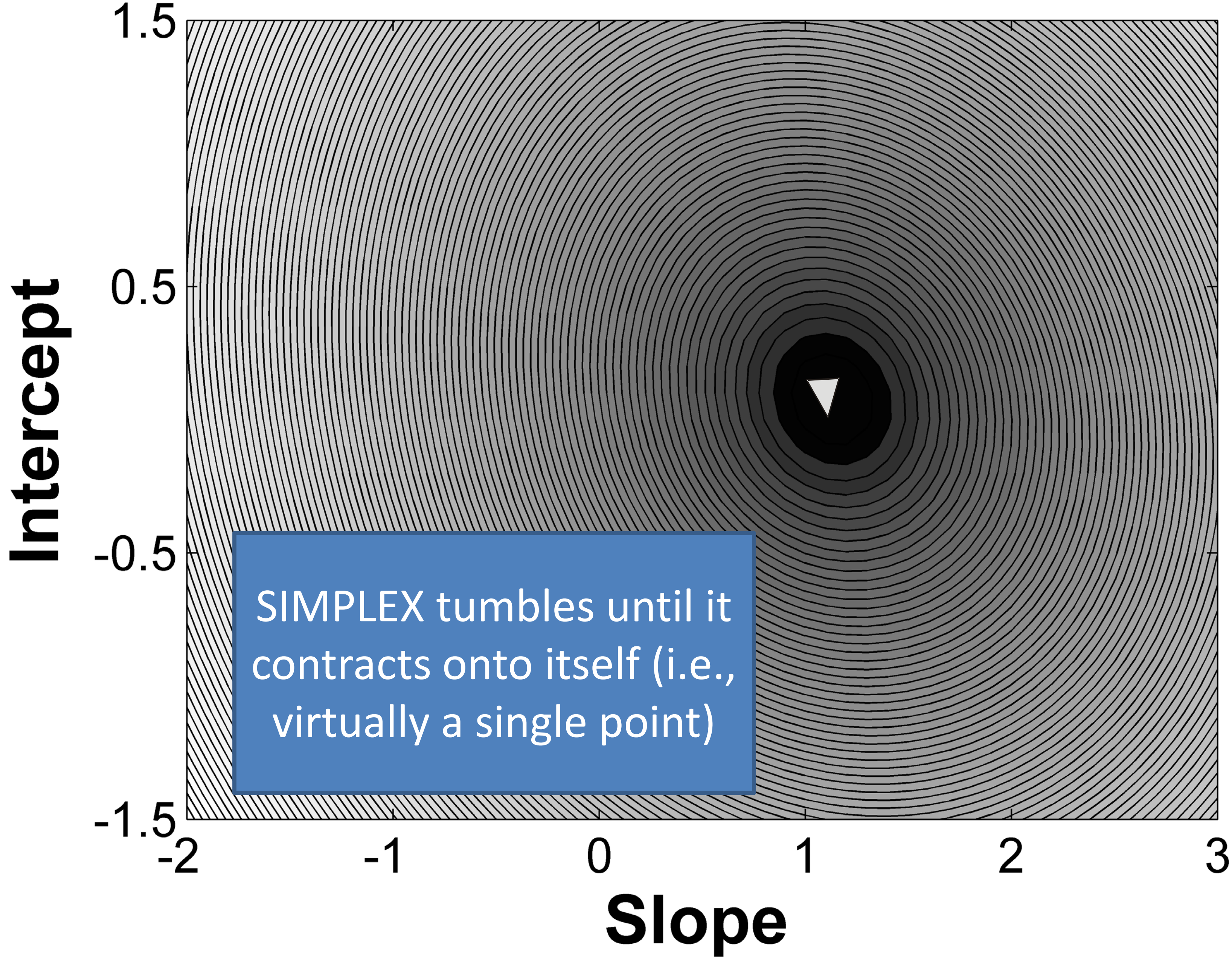
SIMPLEX

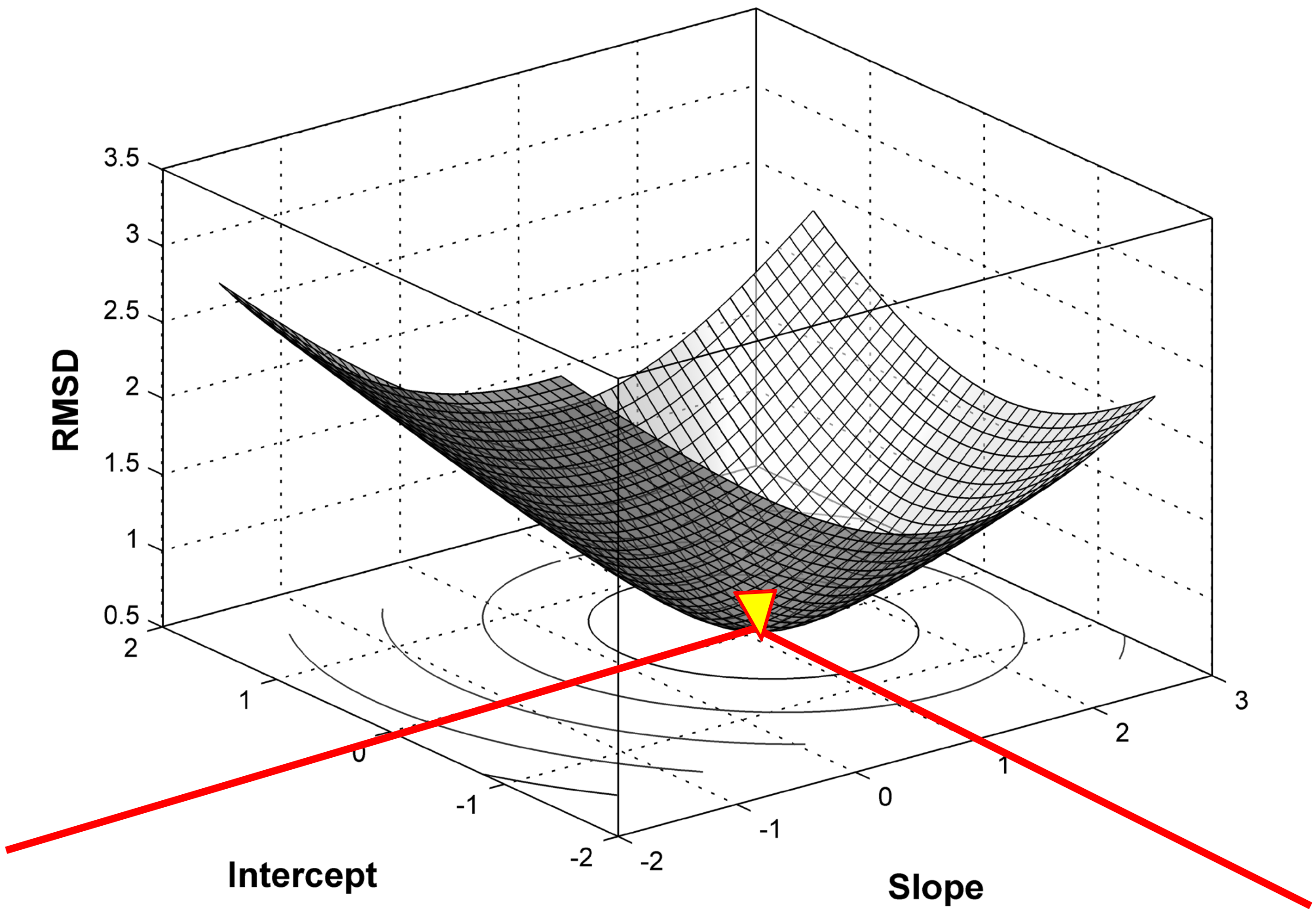
- Once initial simplex has been created
  - by evaluating discrepancy function at each vertex
- It tumbles down error surface
  - somersault towards lowest vertex (“reflection”)
  - if particularly rewarding, may be accompanied by expansion
  - or move points with worst fit closer to center (“contraction”)


1.5
0.5
-0.5
-1.5
-2
-1
0
1
2
3
Slope
Intercept
2D projection of regression error surface



Intercept
1.5
0.5
-0.5
-1.5
-2
-1
0
1
2
3
Slope


1.5
0.5
-0.5
-1.5
-2
-1
0
1
2
3
Intercept
Slope
SIMPLEX tumbles until it contracts onto itself (i.e., virtually a single point)


RMSD
3.5
3
2.5
2
1.5
1
0.5
2
1
0
-1
-2
Intercept
-2
-1
0
1
2
3
Slope

Limitations of SIMPLEX

- Inadvisable to use SIMPLEX with more than 5 parameters (Box, 1966)
  - even 2 can be tough ...
- Discrepancy function must be deterministically related to parameter values
  - random variation turns error surface into "bubbling goo"
  - inevitable with random components (e.g., random-walk simulation)
  - run many replications

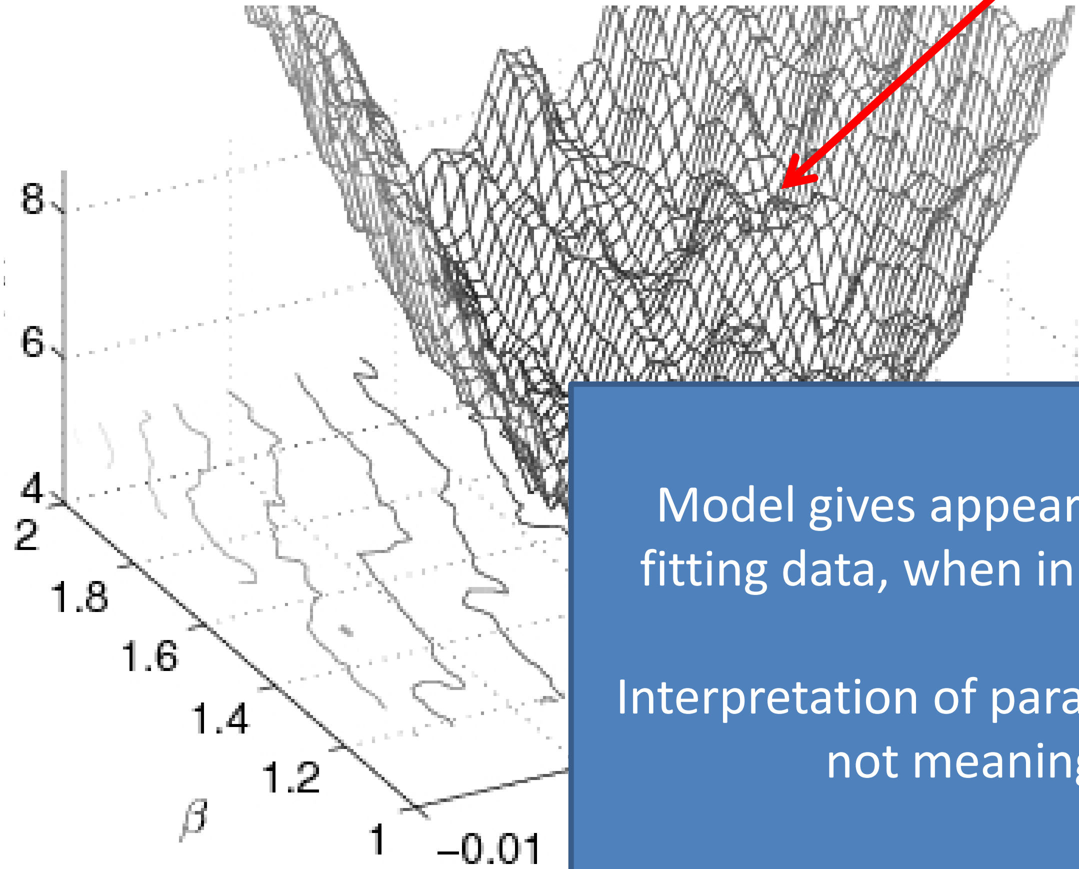
Potential Problems

- SIMPLEX (and other related techniques) can only move downhill
- SIMPLEX is blind to anything but local surroundings
- Hence SIMPLEX may run into trouble when the error surface has a challenging shape
  - trenches
  - ridges
  - plateaus

Local Minima

oops ...

Discrepancy



Model gives appearance of not fitting data, when in fact it could

Interpretation of parameter values not meaningful

Some Possible Solutions

- Multiple starting values
  - convergence to same best-fitting estimates is suggestive of global minimum
  - especially if starting values differ widely
- Simulated annealing
  - SIMPLEX on steroids with a shake
  - can jump out of local minima
  - in Chapter 3 in Farrell & Lewandowsky

SIMPLEX in R