Distributed Optimization for Machine Learning

School of Electrical and Computer Engineering

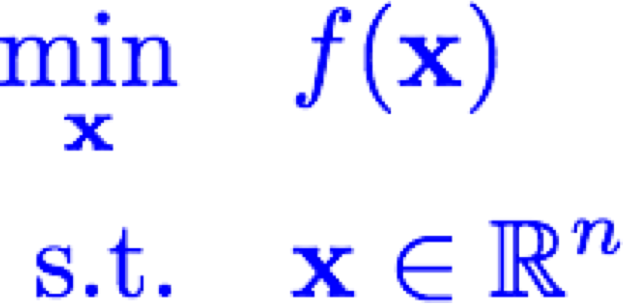
University of Tehran

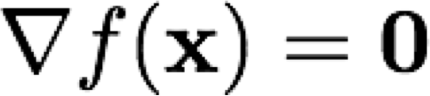
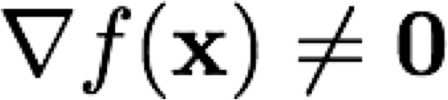
Erfan Darzi

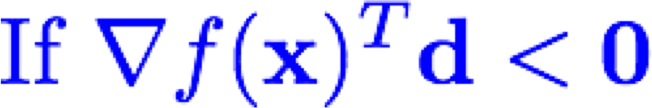
Lecture 3 – Iterative Descent Methods and Convergence Analysis

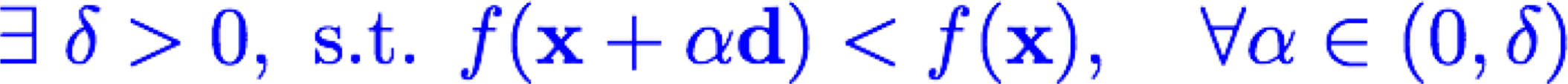
erfandarzi@ut.ac.ir

# Iterative Descent Methods



* If , we have a candidate
* If , not a candidate è Can we locally improve?





How to select

Step-size and direction?

# Choices of Direction

•

Steepest/gradient descent:

•

Diagonally scaled gradient descent:

•

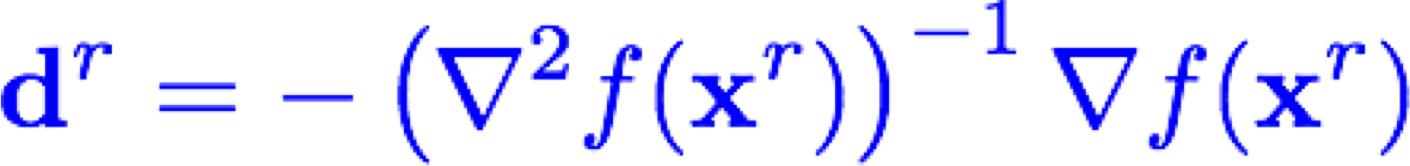
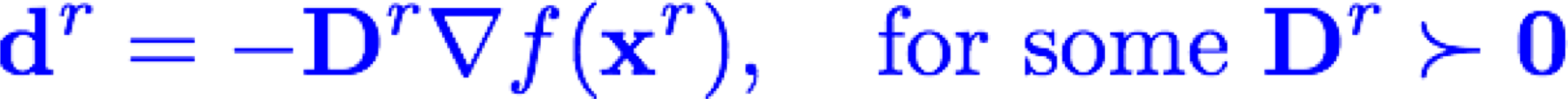
Newton direction (why?):

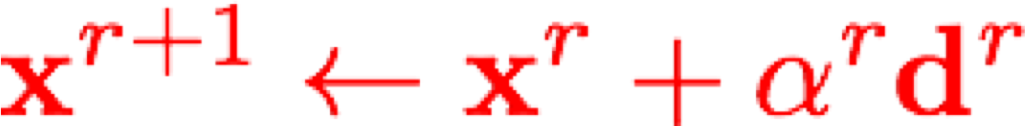
•

Benefit

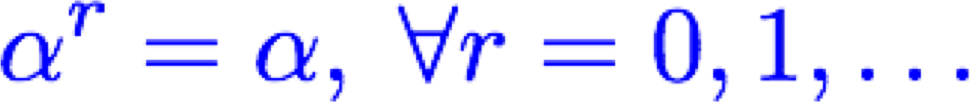
•

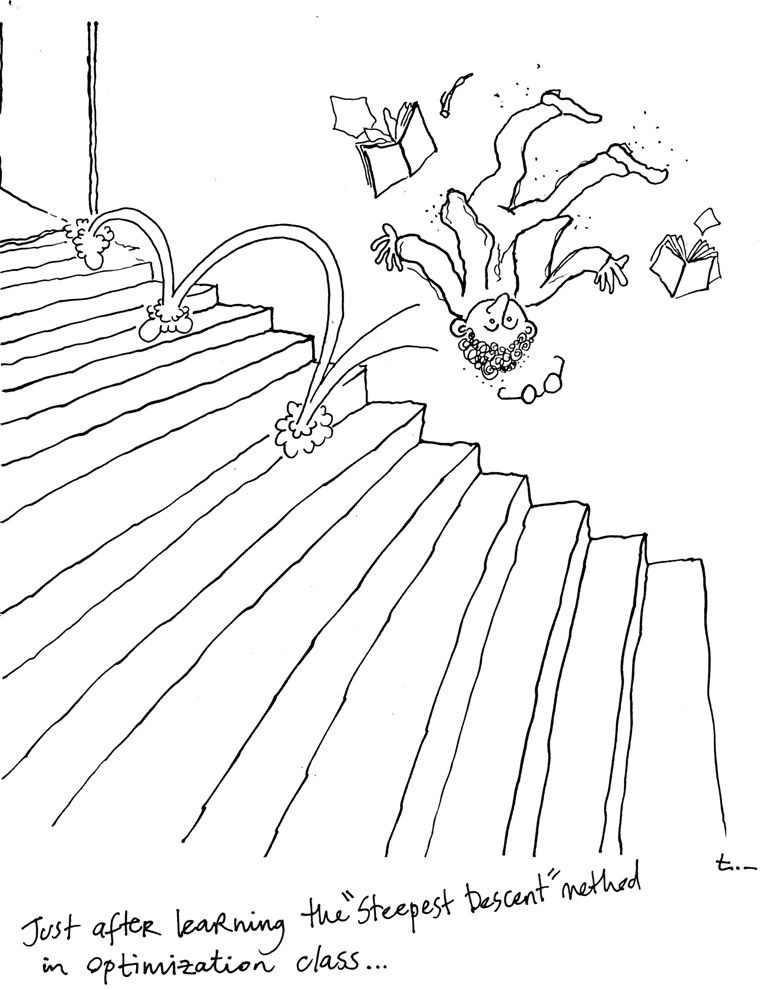
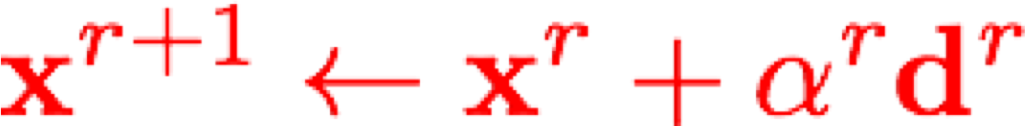
Drawback





Choices of Step-size:

* Constant: 



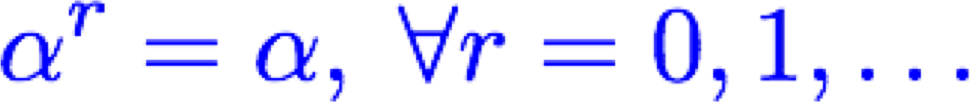
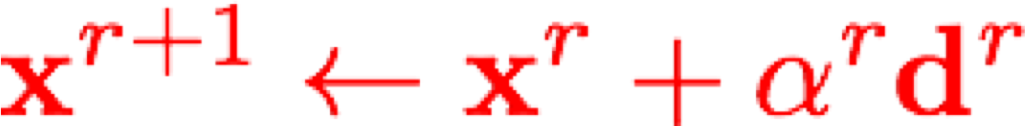
Need to be careful about step-size!!

http://www.eurasip.org/DSPHumour/steepest-descent.jpg

Choices of Step-size:

•

Constant:



•

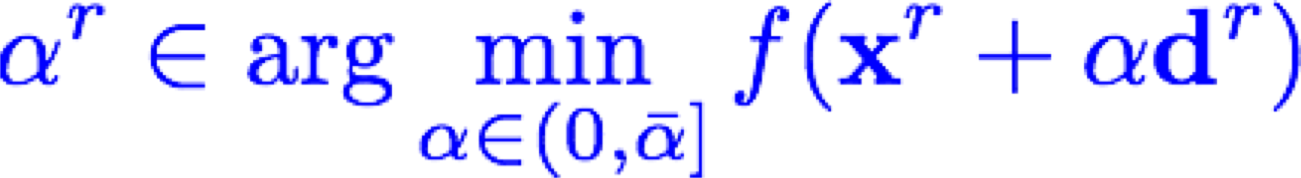
Exact Minimization:



•

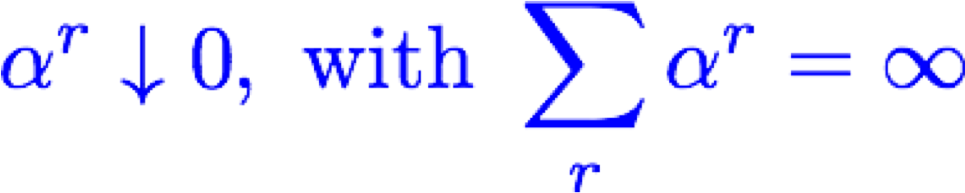
Limited Minimization

:



•

Diminishing:

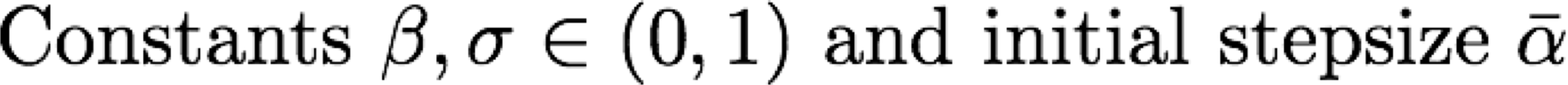


•

Back

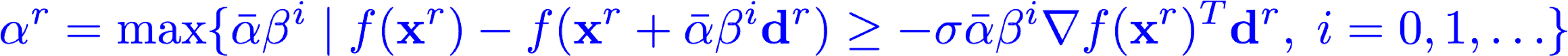
-

tracking/Armijo:



**Predicted decrease**

**Actual decrease**



**Claim**

**:**



Why?

Convergence Analysis 

* Convergence to a stationary point (set of stationary points)
* Typical minimum requirement

•

Asymptotic rate of convergence (

**Convergence rate**

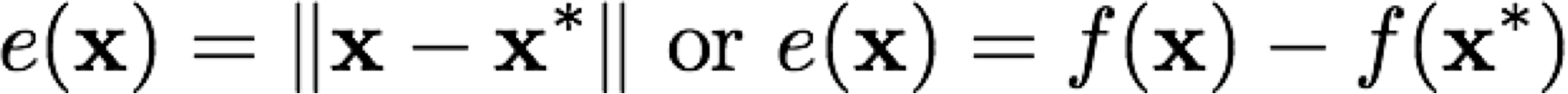
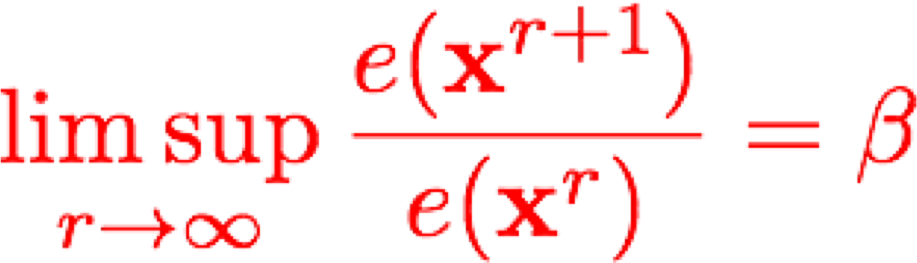
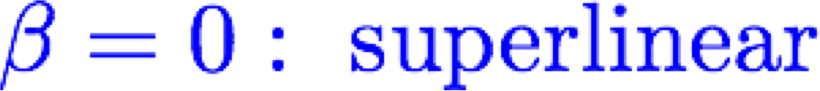
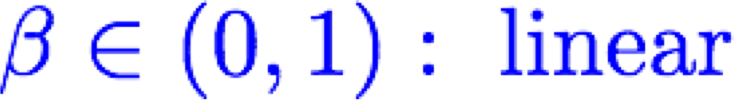
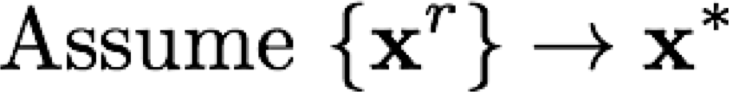
)

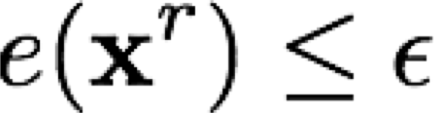
•

Error function examples:

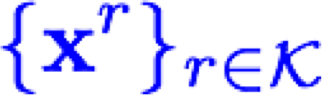
•

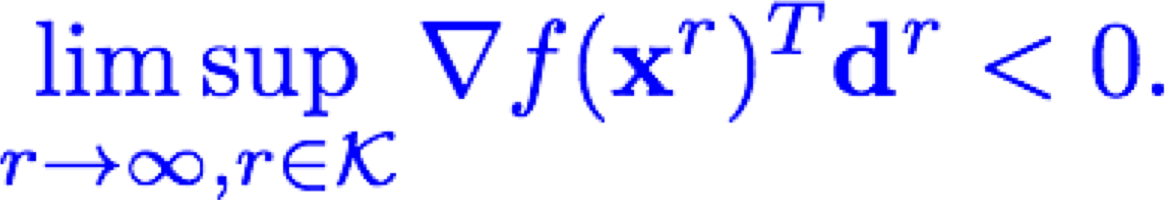
Asymptotic behavior

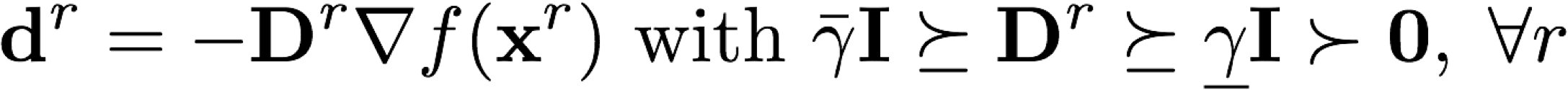


* Iteration complexity analysis: Why we call it linear?
* Number of iterations required to achieve optimal solution: 
* Currently, worst case analysis

# Convergence to Stationary Points

* To a single limit point may not be easy
* **Gradient related condition:** For any subsequence converging to a non-stationary point, the corresponding subsequence is bounded and



* Example: 

# Convergence to Stationary Points

•

Assume:

•

•

gradient related

•

Lipschitz gradient:

•

One of the following

step

-

size rules

:

•

(

a) Diminishing

:

•

(

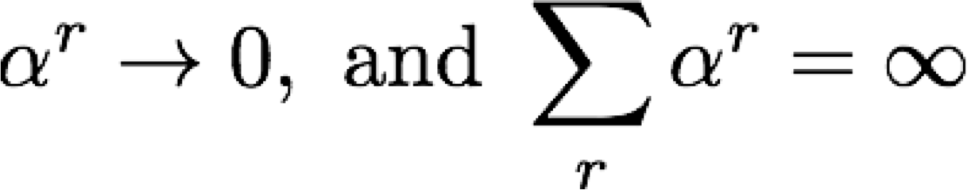
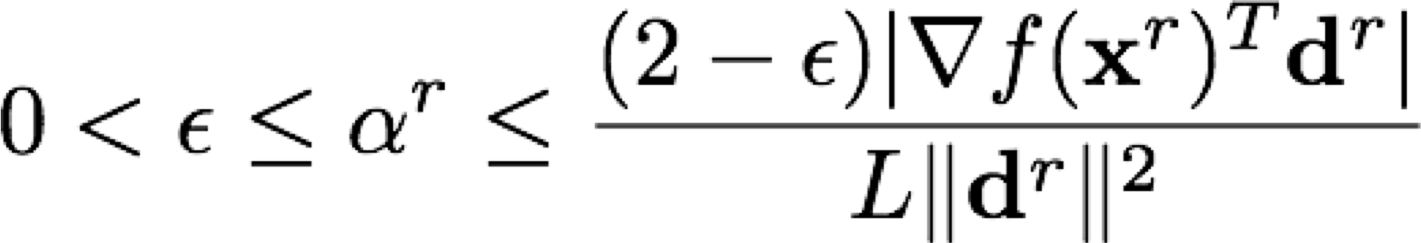
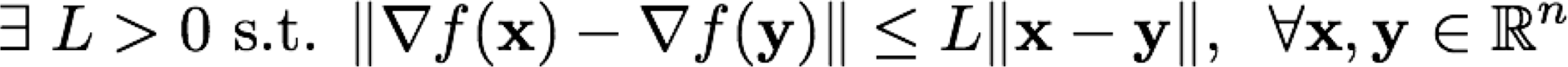
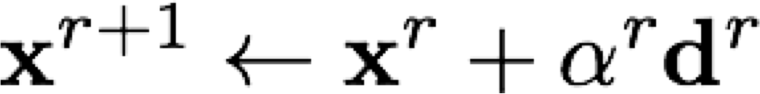
b) Armijo

•

(

c) Small enough

:



•

Then, every limit point of the iterates is a stationary point, i.e.,

•

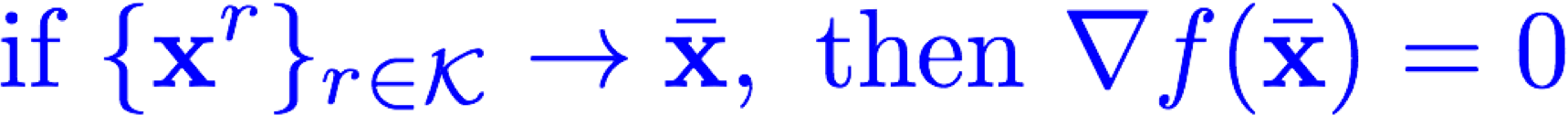
**Special case: gradient direction**

•

**Proof (Requires descent lemma)**

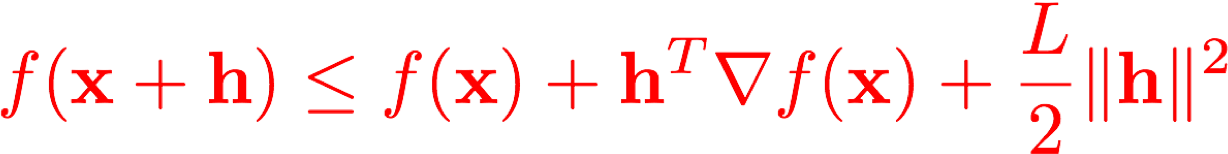
•

**These are (asymptotically) monotone rules**



Why useful?

Proof



No assumption on convexity!

8