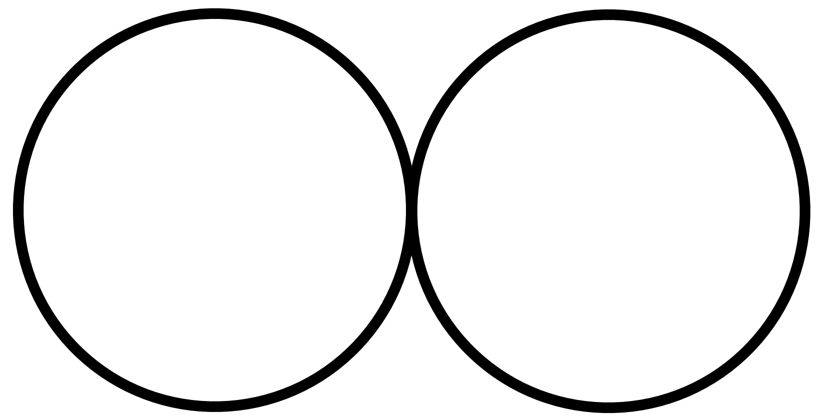
10.34: Numerical Methods Applied to Chemical Engineering

Lecture 10:
Unconstrained Optimization
Steepest decent

Recap

- Homotopy and Bifurcation

Recap



$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} (x_1 - 1)^2 + x_2^2 - 1 \\ (x_1 + 1)^2 + x_2^2 - 1 \end{pmatrix} = 0$$

$$\mathbf{J}(\mathbf{x}) = \begin{pmatrix} 2(x_1 - 1) & 2x_2 \\ 2(x_1 + 1) & 2x_2 \end{pmatrix}$$

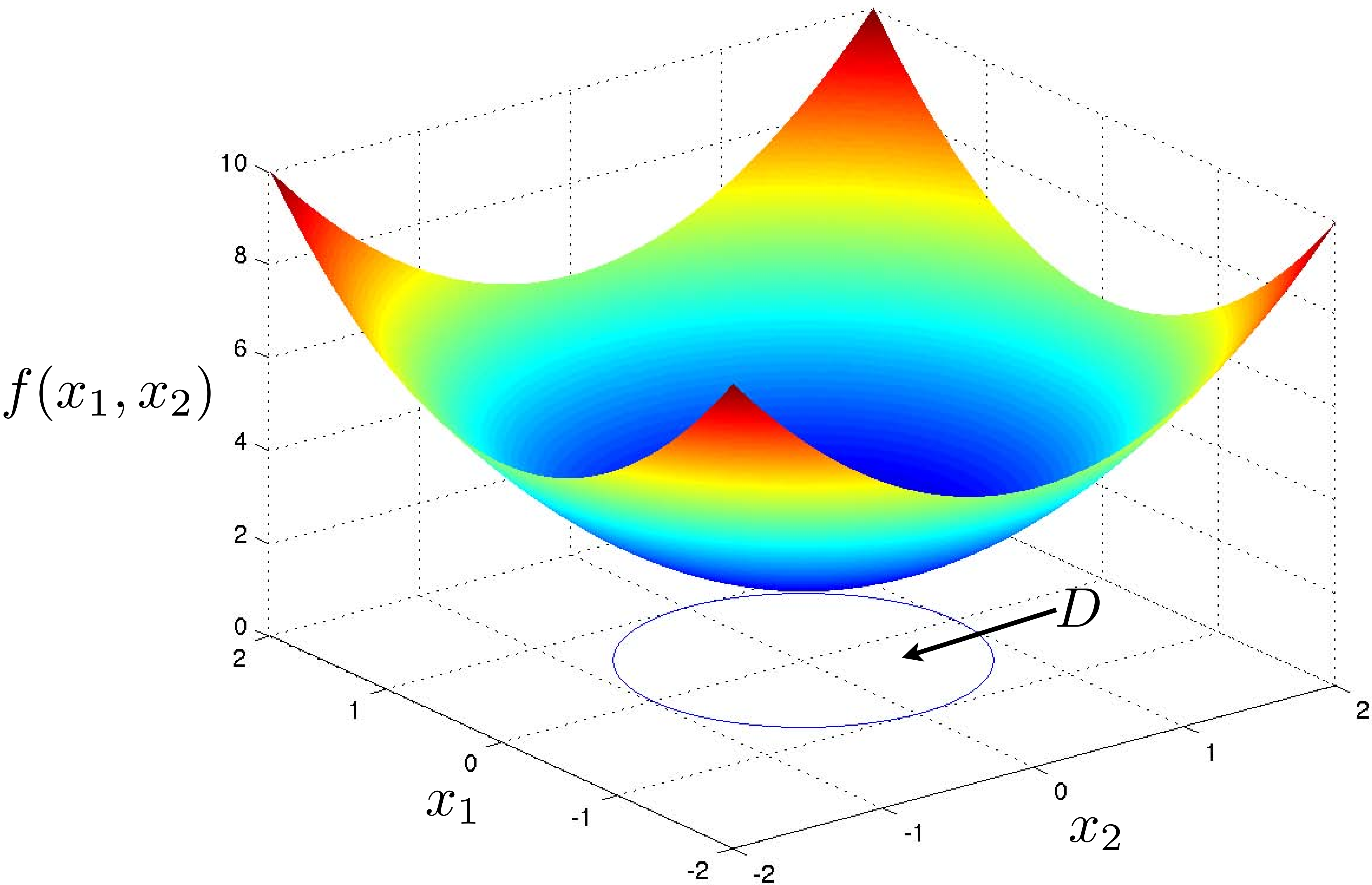
Optimization

- Problems of the sort:

$$\min_{\mathbf{x} \in D} f(\mathbf{x}) \qquad\qquad \arg\min_{\mathbf{x} \in D} f(\mathbf{x})$$

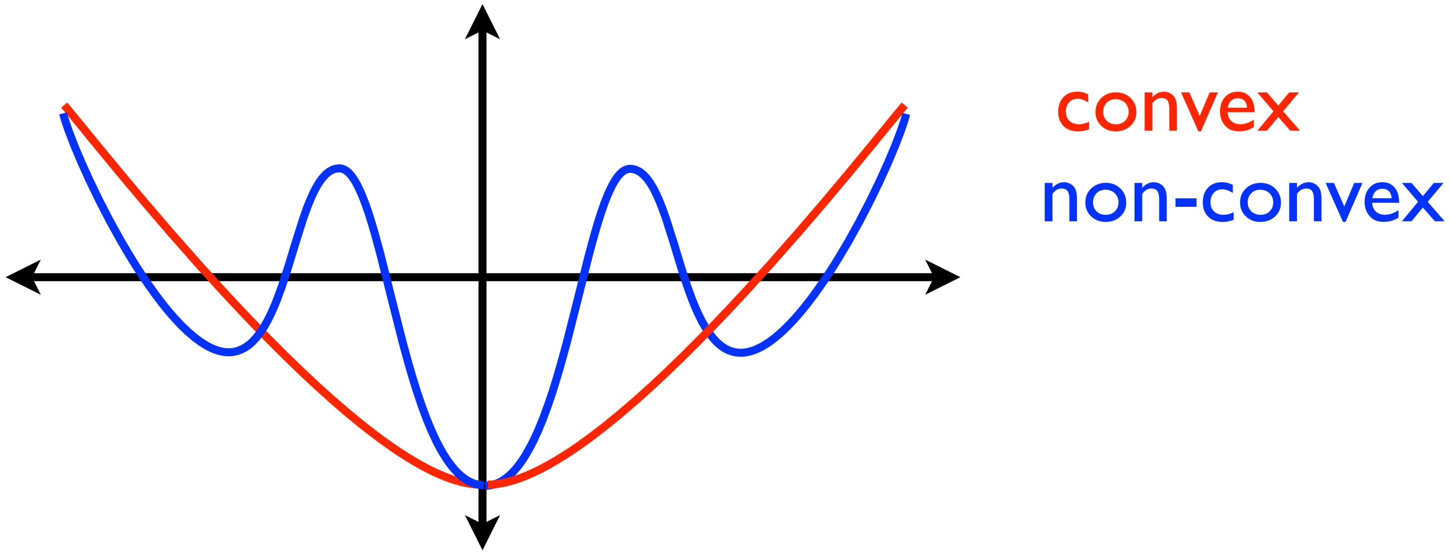
  - $f(\mathbf{x})$: objective function, cost function, energy
    - "metric to compare alternatives"
  - $\mathbf{x}$: "design alternatives"
  - D: feasible set
- Maximization of $f(\mathbf{x})$ is just minimization of $-f(\mathbf{x})$

Optimization



Optimization

- Goal: find $\mathbf{x}^* \in D : f(\mathbf{x}^*) < f(\mathbf{x}) \quad \forall \mathbf{x} \in D$
  - $\mathbf{x}^*$ is not necessarily unique. There could be more than one $\mathbf{x}^*$ in D.
- Convexity: a function is convex if the line connecting any two points above the function is also above the function:



  - Convex functions have a single, global minimum
    - Most algorithms are characterized in terms of their ability to find the global minimum of convex functions.
  - Non-convex function may have global or local minima

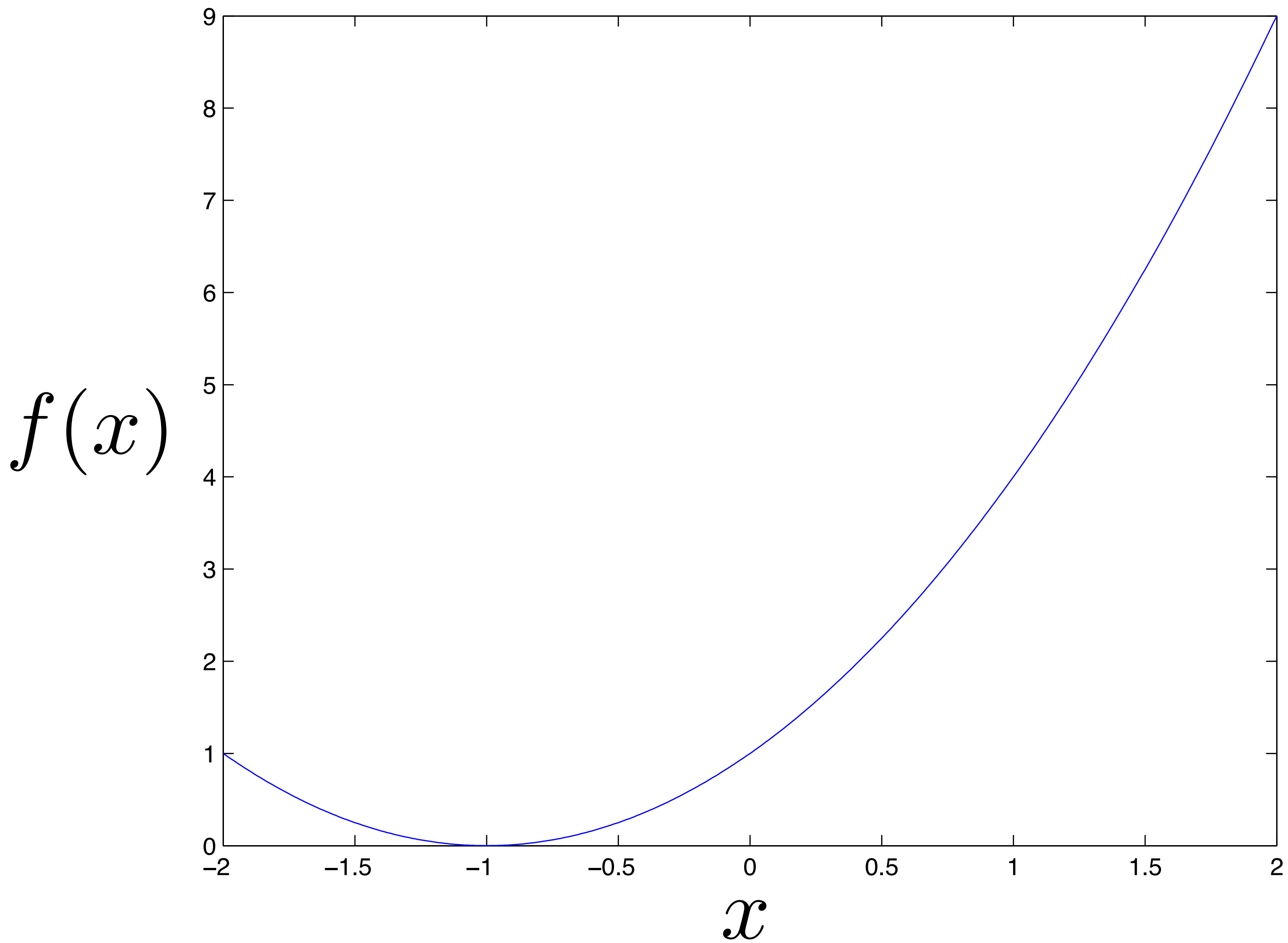
Optimization

- Examples:
  - Find the value of x that minimizes

$$f(x) = x^2 + 2x + 1$$

  - Find the value of $x \in [0, 1]$ that minimizes

$$f(x) = x^2 + 2x + 1$$

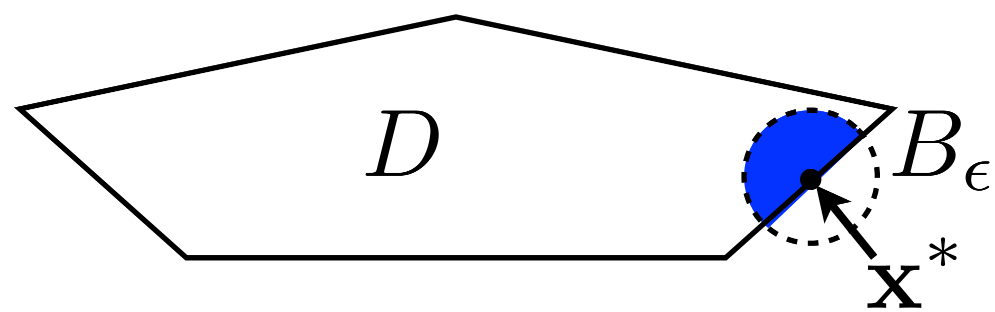


Optimization

- Examples: linear programs
  - Premium and regular ice cream are sold for $5/gallon and $3.5/gallon respectively.
  - Premium ice cream is 30% air by volume while regular ice cream is 50% air by volume.
  - We can produce X gallons of premium and Y gallons of regular ice cream all at the same cost, $1/gallon.
  - What fraction of milk processed should go toward premium versus regular ice cream?

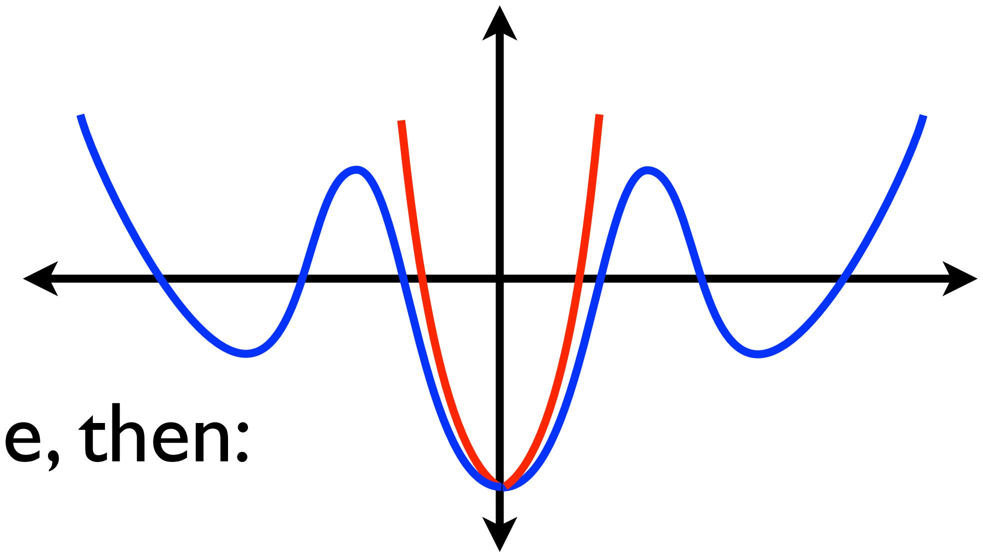
Optimization

- $\mathbf{x}^* \in D$ is a local minimum of
  - if $\exists \quad \epsilon > 0 : f(\mathbf{x}^*) < f(\mathbf{x}), \quad \forall \mathbf{x} \in D \cap B_\epsilon(\mathbf{x}^*)$



  - Global minima are also local minima
  - If $f(\mathbf{x})$ is convex in D then a local minimum is the global minimum in D.
- If D is a closed set, the problem of finding the minimum is called constrained optimization.
- If D is an open set: $\mathbb{R}^N$, the problem of finding the minimum is called unconstrained optimization

Unconstrained Optimization

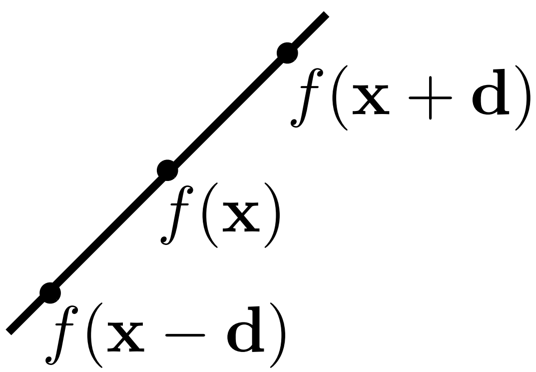


- Optimality criteria:
  - How do I check for local minima?
  - Assume $f(\mathbf{x})$ is twice differentiable, then:

$$f(\mathbf{x}+\mathbf{d}) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x})\mathbf{d} + \ldots$$

  - where: $g_i(\mathbf{x}) = \frac{\partial f}{\partial x_i}$ $\quad H_{ij}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}$
- As $\|\mathbf{d}\|_p \to 0$

$$f(\mathbf{x}+\mathbf{d}) - f(\mathbf{x}) = \mathbf{g}^T\mathbf{d}$$



- If $\mathbf{g}^T\mathbf{d} > 0$, then $f(\mathbf{x}+\mathbf{d}) > f(\mathbf{x})$
  - But, replace $\mathbf{d}$ with $-\mathbf{d}$, and the converse is true
- Therefore, I have a critical point when: $\mathbf{g} = \nabla f(\mathbf{x}) = 0$

Unconstrained Optimization

- Solving unconstrained optimization problems is the same as solving the system of nonlinear equations:

$$\mathbf{g} = \nabla f(\mathbf{x}) = 0$$

  - Except, we want to ensure that we only find the roots associated with local minima in $\mathbf{f}(\mathbf{x})$

$$f(\mathbf{x}+\mathbf{d}) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\mathbf{d} + \frac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x})\mathbf{d} + \ldots$$

    - If the eigenvalues of the Hessian are positive, we can be sure that $\mathbf{f}(\mathbf{x})$ is a minimum. Why?
    - For a minimum, the eigenvalues must be non-negative
  - How do we craft an algorithm that only finds minima?

Unconstrained Optimization

- Examples:
  - Calculate the gradient. Where is the critical point? Calculate the Hessian. Is the critical point a minimum?

$$f(\mathbf{x}) = x_1^2 + x_2^2$$

$$f(\mathbf{x}) = x_1^2 - x_2^2$$

$$f(\mathbf{x}) = x_1^4 + x_2^4$$

Unconstrained Optimization

- Method of steepest decent:
  - Solve the equation: $\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = 0$, iteratively by taking steps in a direction that decreases $f(\mathbf{x})$

    $$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{d}_i$$

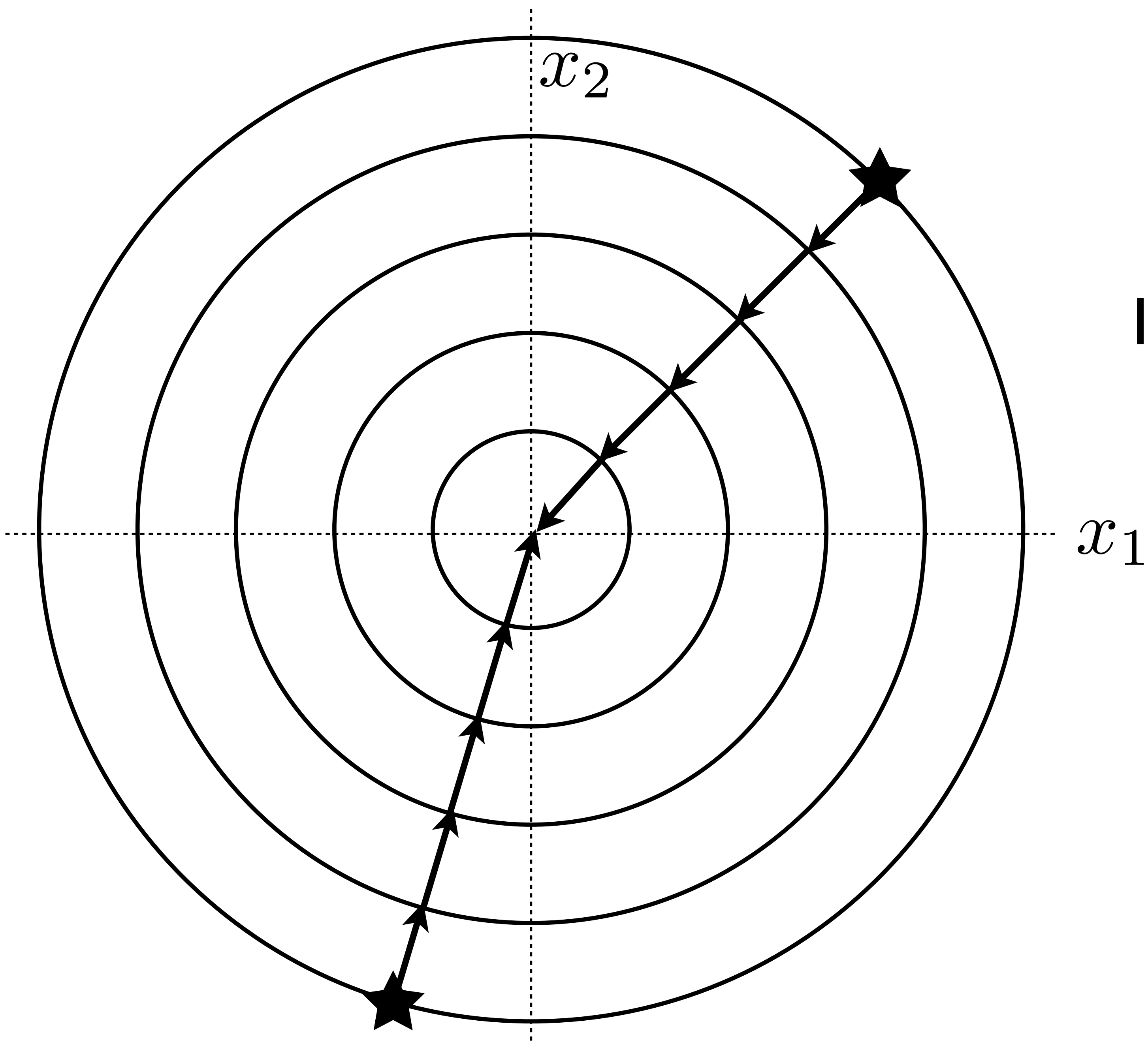
    - with $\alpha_i > 0$ and $\mathbf{g}(\mathbf{x}_i)^T \mathbf{d}_i < 0$
    - This ensures that $\mathbf{d}_i$ is a descent direction:

      $$f(\mathbf{x}_i + \alpha_i \mathbf{d}_i) = f(\mathbf{x}_i) + \alpha_i \mathbf{g}(\mathbf{x}_i)^T \mathbf{d}_i + \ldots$$

    - Which descent direction should I choose?
      - One option: maximize $-\mathbf{g}(\mathbf{x}_i)^T \mathbf{d}_i$
        - C-S inequality: $-\mathbf{g}(\mathbf{x}_i)^T \mathbf{d}_i \leq \|\mathbf{g}(\mathbf{x}_i)\|_2 \|\mathbf{d}_i\|_2$
        - Solution: let $\mathbf{d}_i = -\mathbf{g}(\mathbf{x}_i)$

Unconstrained Optimization

- Method of steepest decent:
  - Example: $f(\mathbf{x}) = x_1^2 + x_2^2$
    - Contours for the function:



$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \mathbf{g}(\mathbf{x}_i)$$

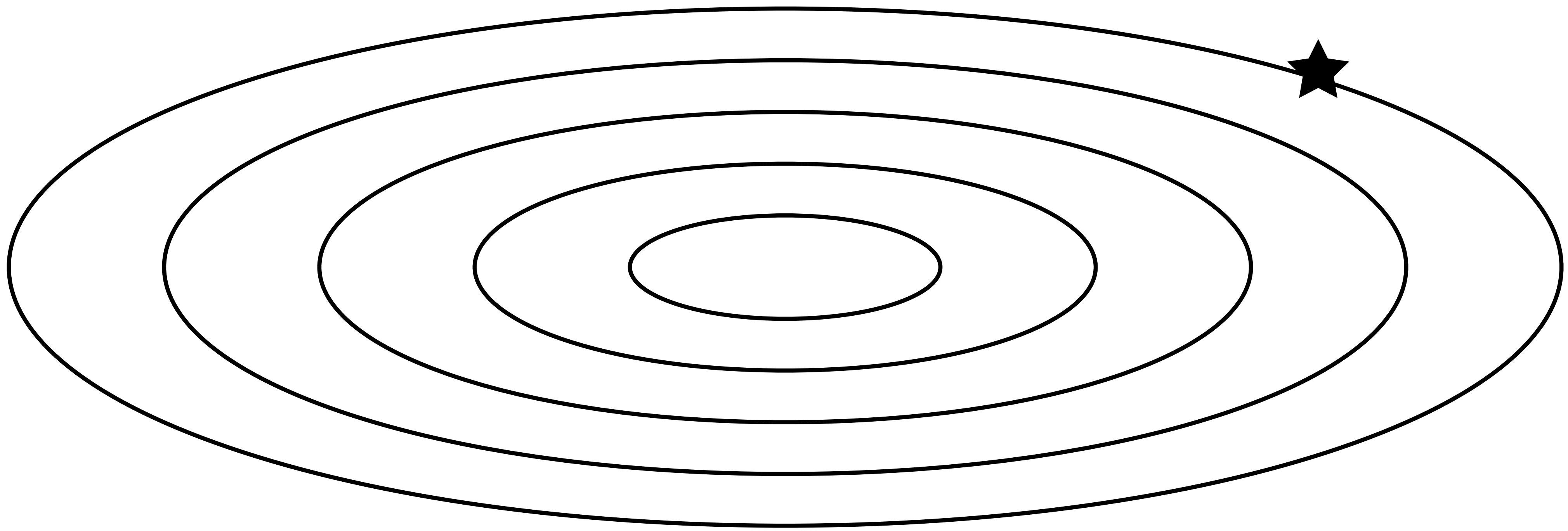
Is there a best value of α_i to use with this function?

Unconstrained Optimization

- Method of steepest decent:
  - Direction of steepest descent: $\mathbf{d}_i = -\mathbf{g}(\mathbf{x}_i)$
  - Iterative solution: $\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \mathbf{g}(\mathbf{x}_i)$
    - For small, positive values of α_i, the iterates continue to reduce $f(\mathbf{x})$ until $\mathbf{g}(\mathbf{x}) = 0$
    - The iterative method converges to local minima and potentially saddle points. Need to check the Hessian still to be sure of minima.
  - How do I choose values for α_i?
    - Ideally, we pick the α_i that leads to the smallest value of $f(\mathbf{x}_{i+1})$, but this is its own optimization.
      - We can approximate the solution with a line search like in damped Newton-Raphson.

Unconstrained Optimization

- Method of steepest decent:
  - Example: $f(\mathbf{x}) = x_1^2 + 10x_2^2$
    - Contours for the function:

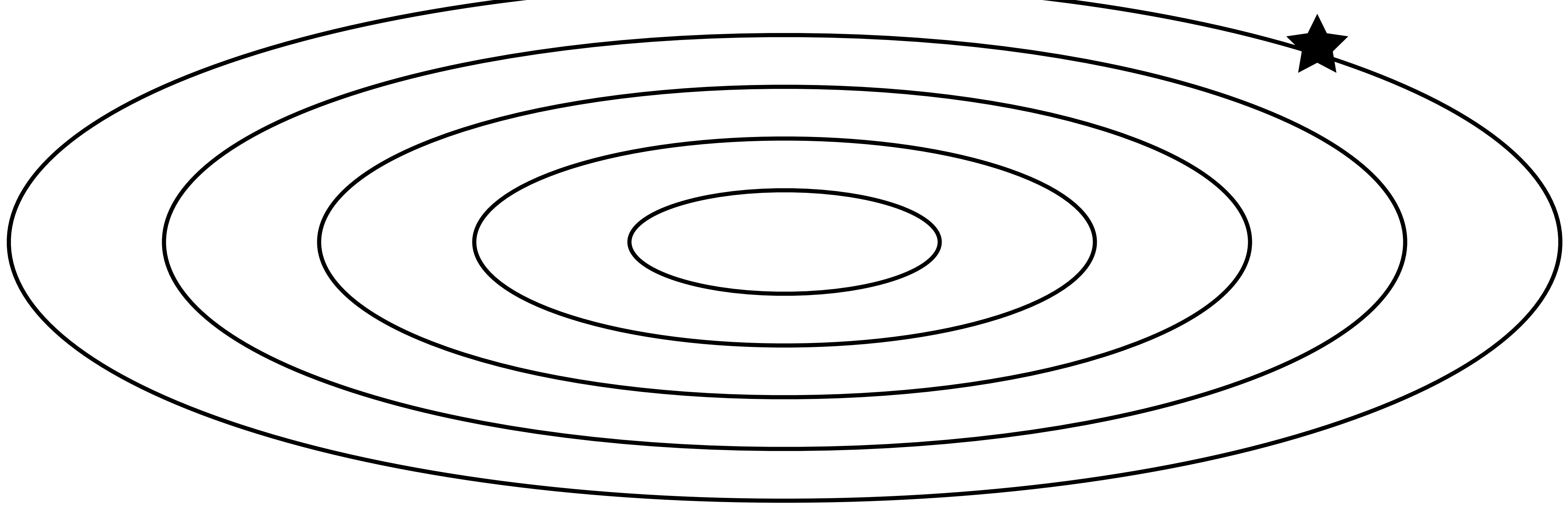


Draw the path given by small α_i

- The choice of α_i is critical!
  - Too small and the convergence is slow

Unconstrained Optimization

- Method of steepest decent:
  - Example: $f(\mathbf{x}) = x_1^2 + 10x_2^2$
    - Contours for the function:

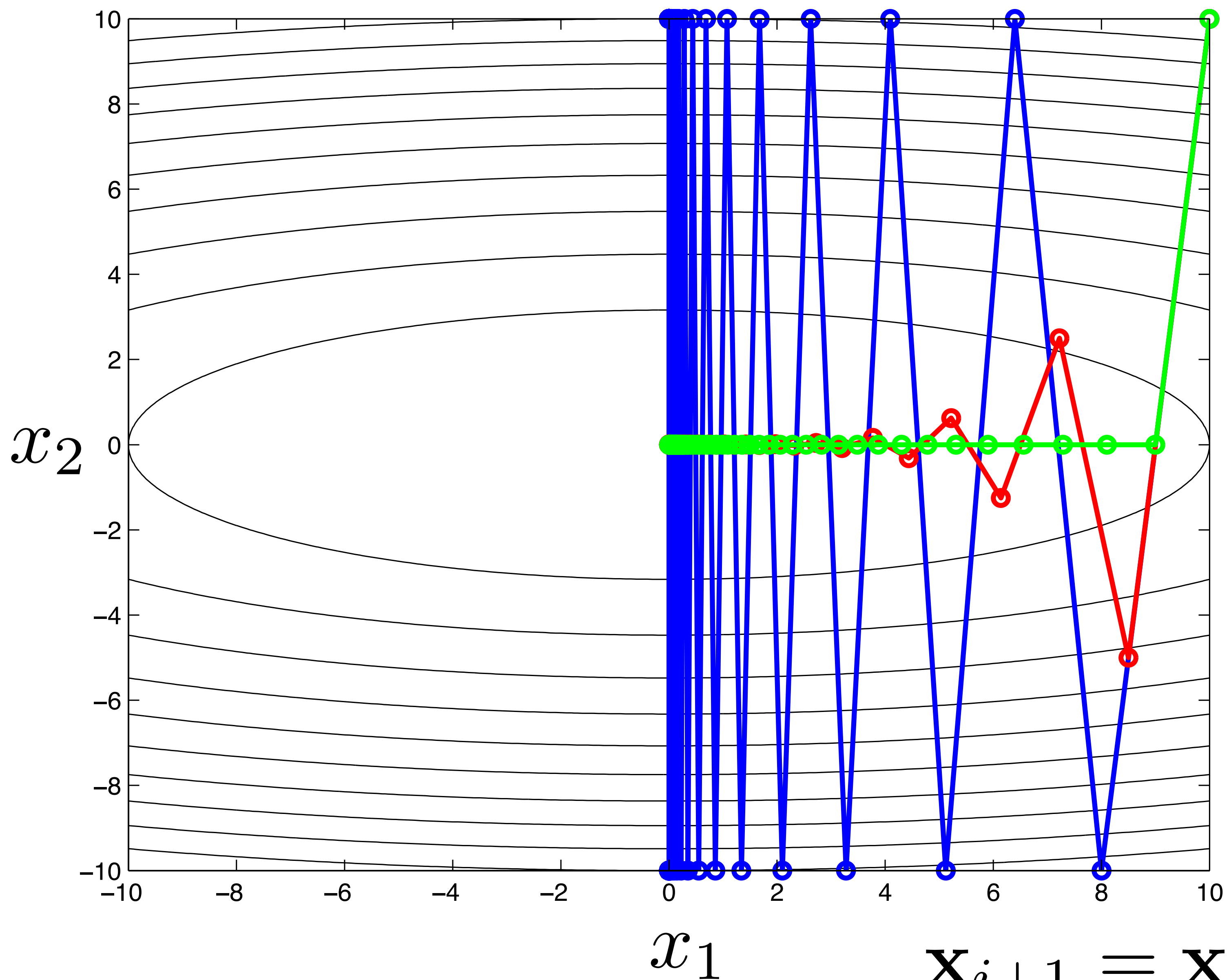


Draw the path given by larger α_i

- The choice of α_i is critical!
  - Too big and convergence is erratic

Unconstrained Optimization

- Method of steepest decent:
  - Example: $f(\mathbf{x}) = x_1^2 + 10x_2^2$
    - Contours for the function:



$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \mathbf{g}(\mathbf{x}_i)$$

Unconstrained Optimization

- Method of steepest decent:
  - Estimating an optimal α_i:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \mathbf{g}(\mathbf{x}_i)$$

    - Use a Taylor expansion:

$$f(\mathbf{x}_{i+1}) = f(\mathbf{x}_i) - \alpha_i \mathbf{g}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_i) + \frac{1}{2}\alpha_i^2 \mathbf{g}(\mathbf{x}_i)^T \mathbf{H}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_i) + \ldots$$

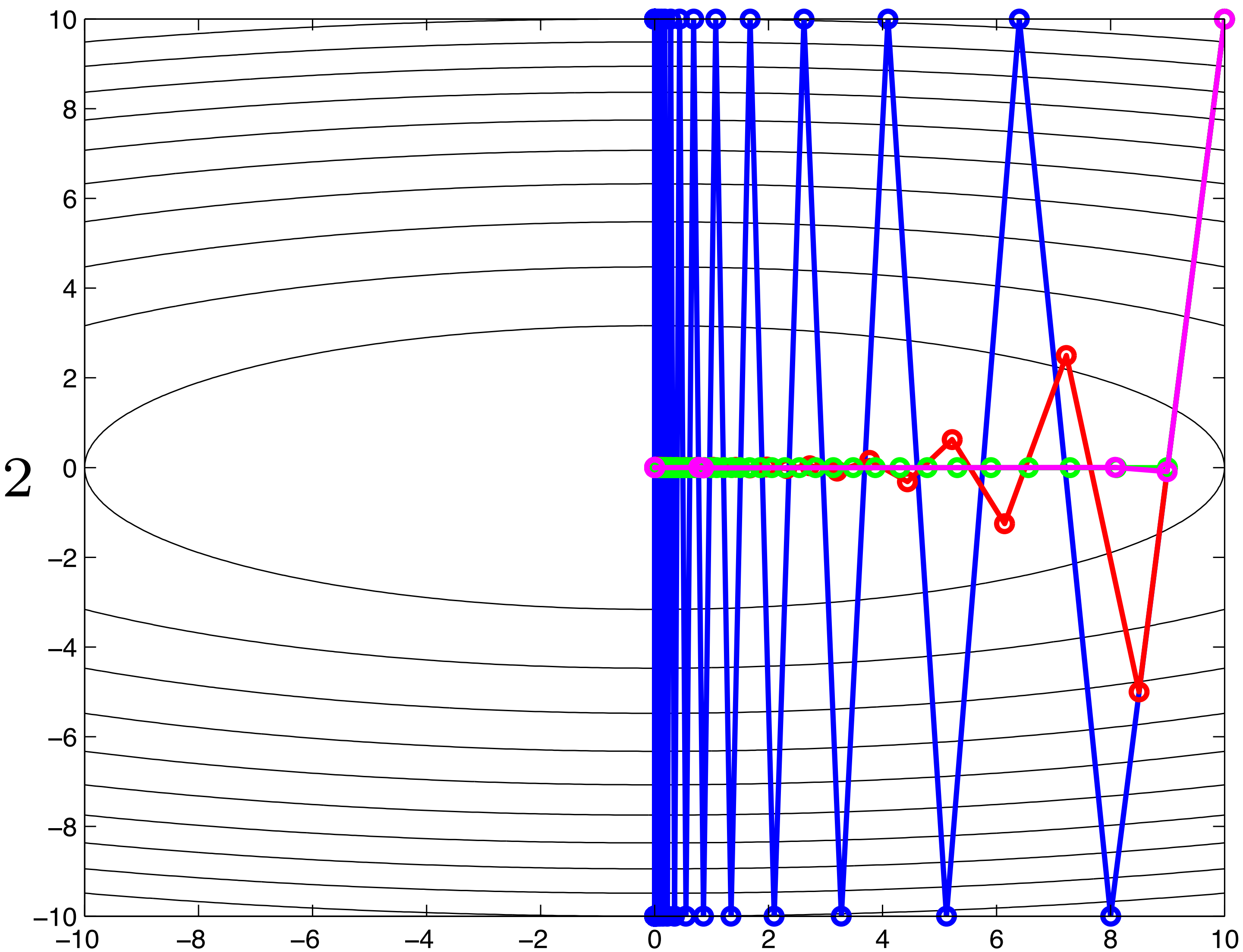
  - This is quadratic in α_i, so find the minimum:

$$\alpha_i = \frac{\mathbf{g}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_i)}{\mathbf{g}(\mathbf{x}_i)^T \mathbf{H}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_i)}$$

    - This can serve as a good starting point for a backtracking line search.

Unconstrained Optimization

- Method of steepest decent:
  - Example: $f(\mathbf{x}) = x_1^2 + 10x_2^2$
    - Contours for the function:



quadratic approximation:

$$\alpha_i = \frac{\mathbf{g}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_i)}{\mathbf{g}(\mathbf{x}_i)^T \mathbf{H}(\mathbf{x}_i)\mathbf{g}(\mathbf{x}_i)}$$

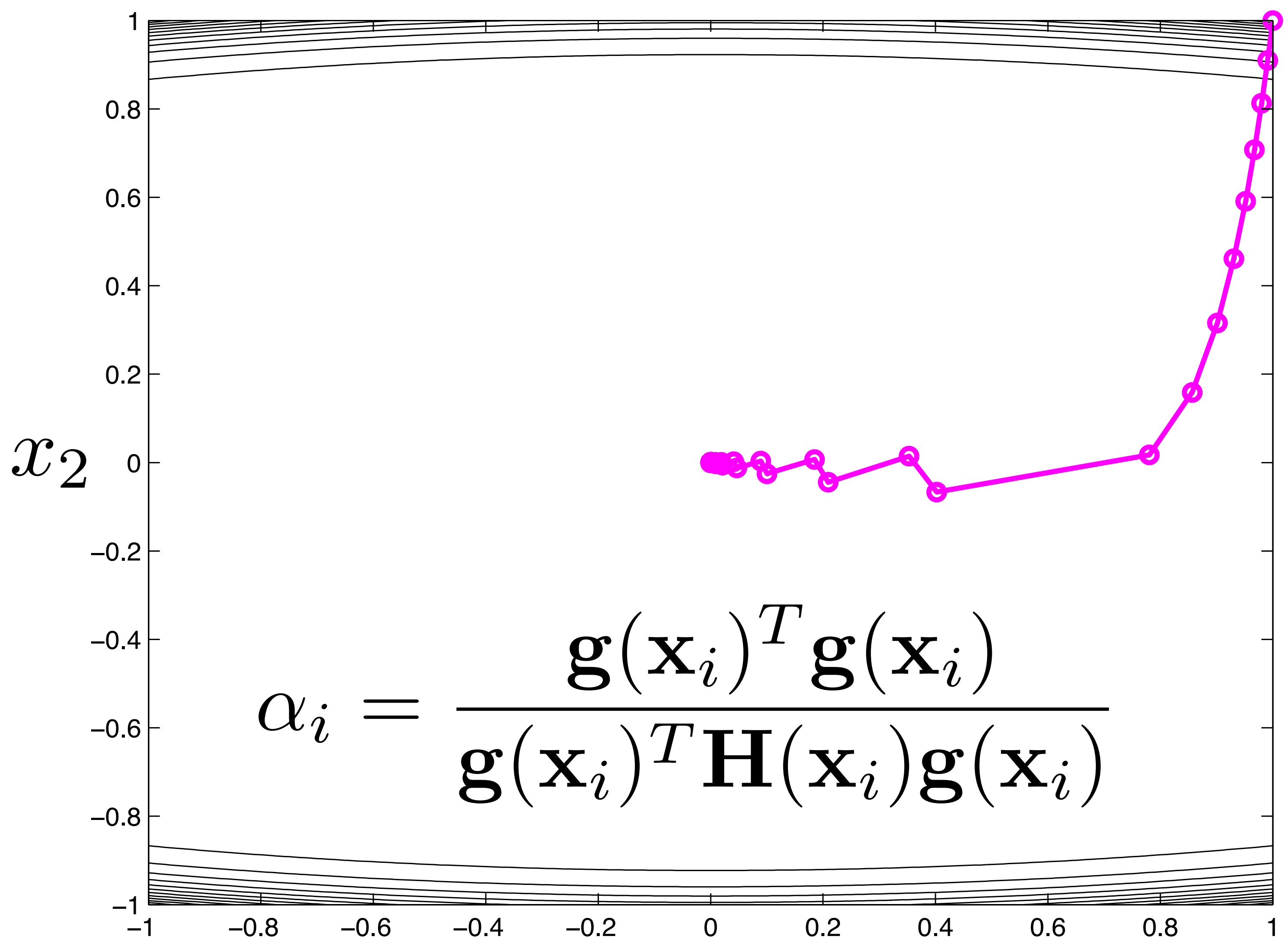
x_2

x_1

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \mathbf{g}(\mathbf{x}_i)$$

Unconstrained Optimization

- Method of steepest decent:
  - Example: $\log f(\mathbf{x}) = x_1^2 + 10x_2^2$
    - Contours for the function:



$$\alpha_i = \frac{\mathbf{g}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_i)}{\mathbf{g}(\mathbf{x}_i)^T \mathbf{H}(\mathbf{x}_i) \mathbf{g}(\mathbf{x}_i)}$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \mathbf{g}(\mathbf{x}_i)$$

10.34 Numerical Methods Applied to Chemical Engineering
Fall 2015