Convex Optimization Refresher

CS 584: Big Data Analytics

Material adapted from
John Duchi (https://www.cs.berkeley.edu/~jordan/courses/294-fall09/lectures/optimization/slides.pdf)
& Stephen Boyd (https://web.stanford.edu/class/ee364a)

Optimization Problem

- Minimize a function subject to some constraints

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_k(x) \le 0, k = 1, 2, \cdots, K \\ & h_j(x) = 0, j = 1, 2, \cdots, J \end{aligned}$$

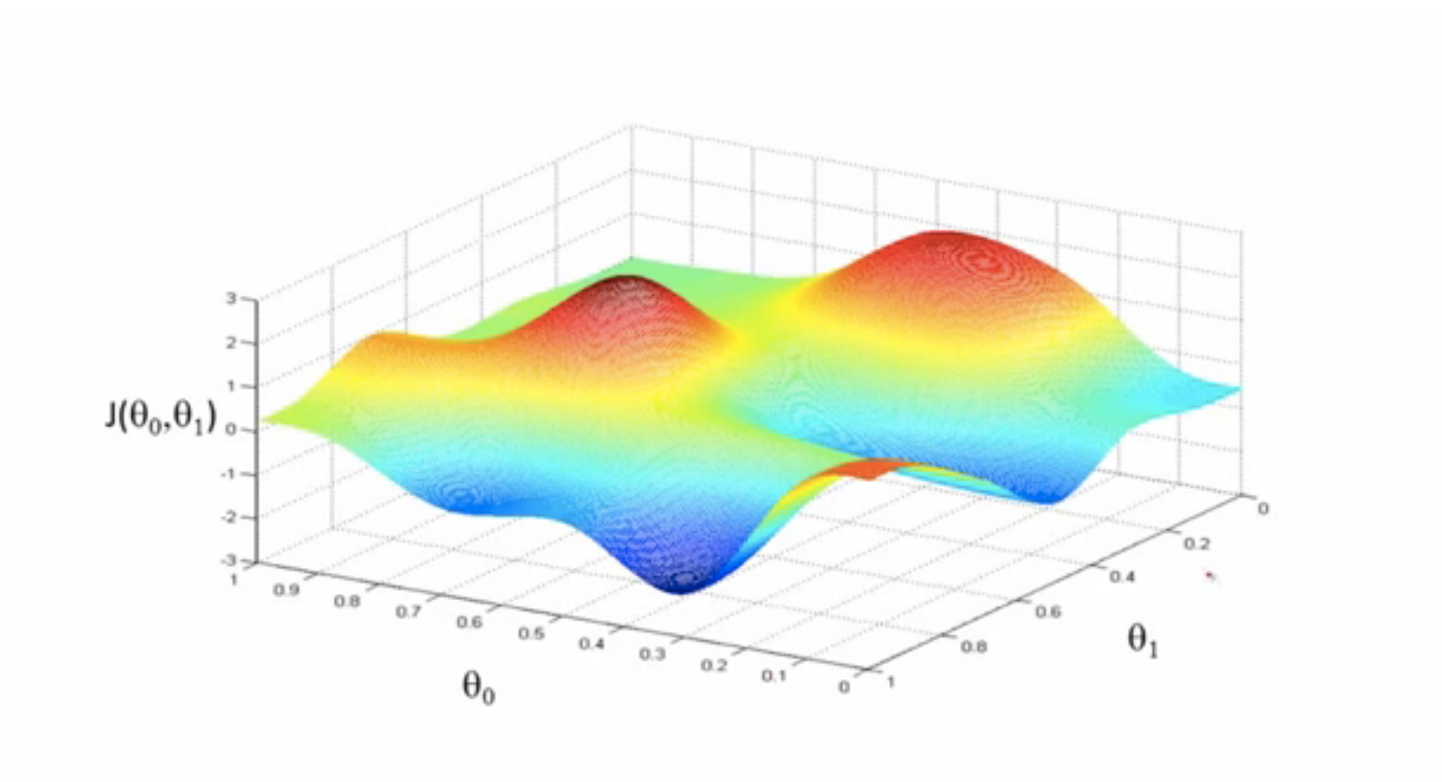
- Example: Minimize the variance of your returns while earning at least \$100 in the stock market.

Machine Learning and Optimization

- Linear regression $$\min_w \; ||Xw - y||^2$$
- Logistic regression $$\min_w \; \sum_i \log(1 + \exp(-y_i x_i^\top w))$$
- SVM $$\min_w \; ||w||^2 + C \sum_i \xi_i$$ $$\text{s.t. } \xi_i \geq 1 - y_i x_i^\top w$$ $$\xi_i \geq 0$$
- And many more ...

Non-Convex Problems are Everywhere

- Local (non-global) minima
- All kinds of constraints



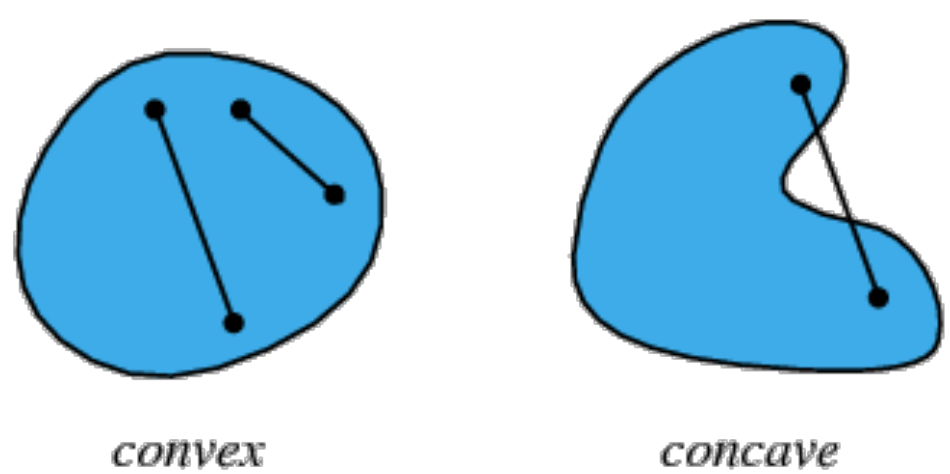
No easy solution
for these problems

Consider
convex problems

Convex Sets

A set C is convex such that given any two points a, b in that set, the line segment between the two points is in the set

$$x_1, x_2 \in C, 0 \leq \theta \leq 1 \implies \theta x_1 + (1-\theta) x_2 \in C$$



Examples: Convex Set

- Real space: $\mathbb{R}^n$
- Non-negative orthant: $\mathbb{R}^n_+$
- Norm balls: $\{x \mid ||x - x_c|| \leq r\}$
- Hyperplane: $\{x \mid a^\top x = b\}, a \neq 0$
- Halfspace: $\{x \mid a^\top x \leq b\}, a \neq 0$

Convexity-preserving operations

- Intersection
- Affine functions $f(x) = Ax + b,\ A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m$
  (e.g., scaling, translation, projection)
- Perspective function
  $$P(x, t) = \frac{x}{t},\ \mathbf{dom}\ P = \{(x, t) \mid t > 0\}$$
- Linear-fractional functions
  $$f(x) = \frac{Ax + b}{c^\top x + d},\ \mathbf{dom}\ f = \{x \mid c^\top x + d > 0\}$$

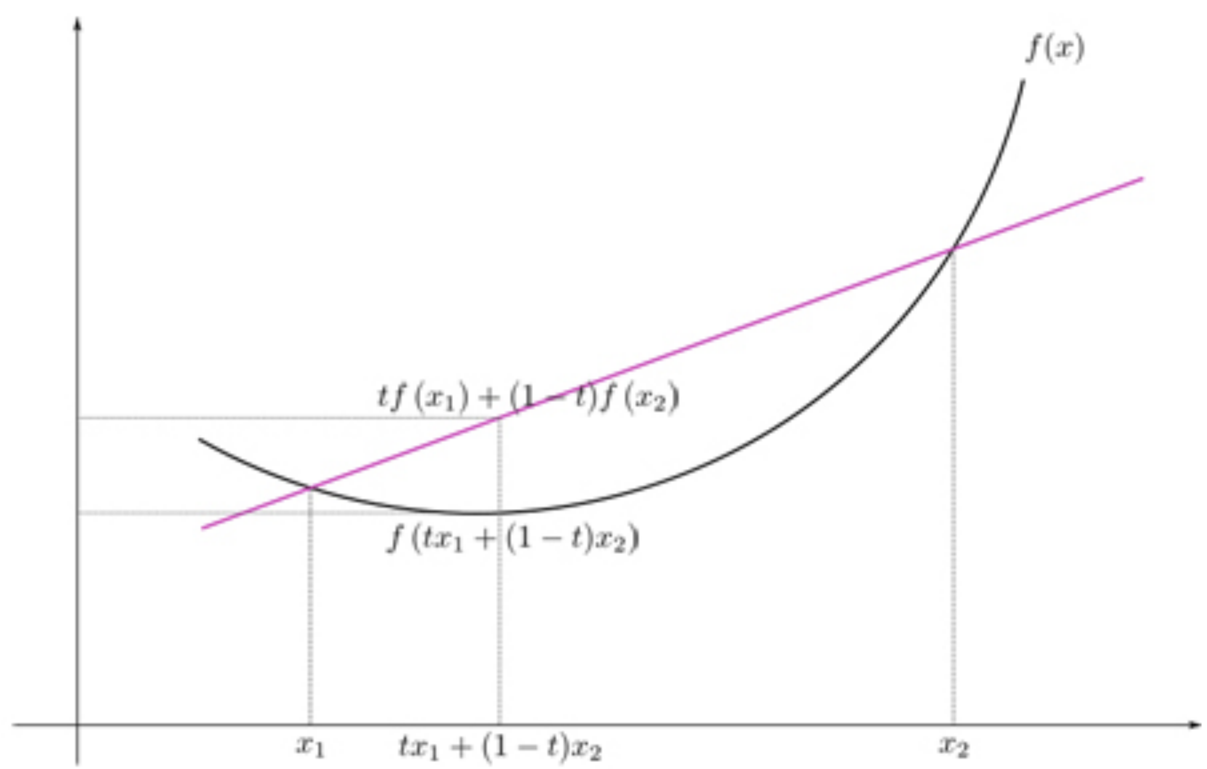
Convex Functions

Definition

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\mathbf{dom}\ f$ is a convex set and

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

for all $x, y \in \mathbf{dom}\ f, 0 \leq \theta \leq 1$



Examples: Convex Functions (Real space)

- affine: $ax + b, \text{for any } a, b \in \mathbb{R}$
- exponential: $e^{ax}, \text{for any } a \in \mathbb{R}$
- powers: $x^\alpha, \;\text{for } \alpha \geq 1 \text{ or } \alpha \leq 0, \; x \in \mathbb{R}_{++}$
- powers of absolute value: $|x|^p, \;\text{for } p \geq 1$
- negative entropy: $x \log x, \; x \in \mathbb{R}_{++}$

Convex Optimization Problem

Definition:
An optimization problem is **convex** if its objective is a convex function, the inequality constraints are convex, and the equality constraints are affine

$$\min_{x} \; f_0(x)$$ convex function

$$\text{s.t. } f_k(x) \leq 0, k = 1, 2, \cdots, K$$ convex sets

$$h_j(x) = 0, j = 1, 2, \cdots, J$$ affine constraints

Benefits of Convexity

- Theorem: If x is a local minimizer of a convex optimization problem, it is a **global** minimizer

- Theorem: If the gradient at c is zero, then c is the global minimum of f(x)

$$\nabla f(c) = 0 \iff c = x^*$$

Lagrange Duality

- Bound or solve an optimization problem via a different optimization problem
- Reformulate the problem as an augment objective with a weighted sum of constraints
  - Remove constraints
  - Introduce new variables
  - Form a dual function

Constructing the dual

Original optimization problem

$$\min_x \; f_0(x)$$

$$\text{s.t. } f_k(x) \leq 0, k = 1, 2, \cdots, K$$

$$h_j(x) = 0, j = 1, 2, \cdots, J$$

dual function

$$g(\lambda, v) = \inf_x \; \{f_0(x) + \sum_k \lambda_k f_k(x) + \sum_j v_j h_j(x)\}$$

$$\lambda_i \geq 0, v_i \in \mathbb{R}$$

Two Properties of Dual

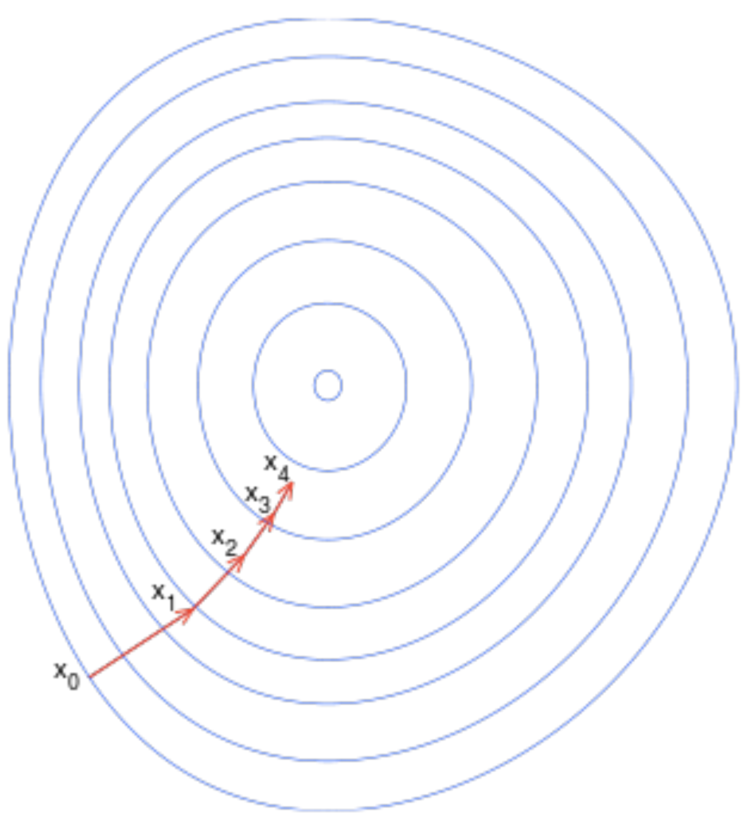
- **Weak Duality** (Lemma): $\text{If } \lambda \geq 0, \text{ then } g(\lambda, v) \leq f_0(x^*)$
  - Always holds for convex and non convex problems
  - Can be used to find nontrivial lower bound for difficult problems
- **Strong Duality** (Theorem): $d^* = x^*$
  - (Usually) holds for convex problems
  - Constraint qualifications are conditions that guarantee strong duality in convex problems

Unconstrained Optimization Algorithms

$$\min_x f(x)$$

Gradient Descent (Steepest Descent)

- Simplest and extremely popular
- Main Idea: take a step proportional to the negative of the gradient
- Easy to implement
- Each iteration is relatively cheap
- Can be slow to converge



Gradient Descent Algorithm

Algorithm 1: Gradient Descent

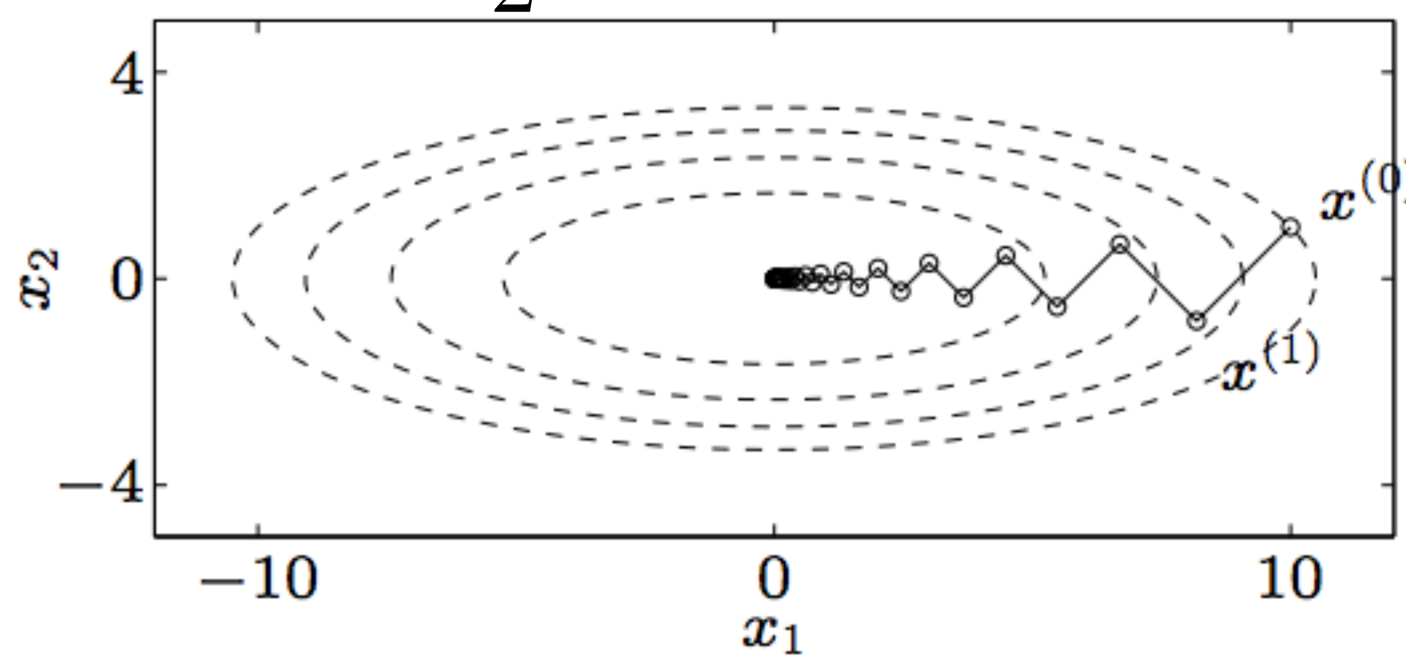
while *Not Converged* **do**

$x^{(k+1)} = x^{(k)} - \eta^{(k)} \nabla f(x)$

end

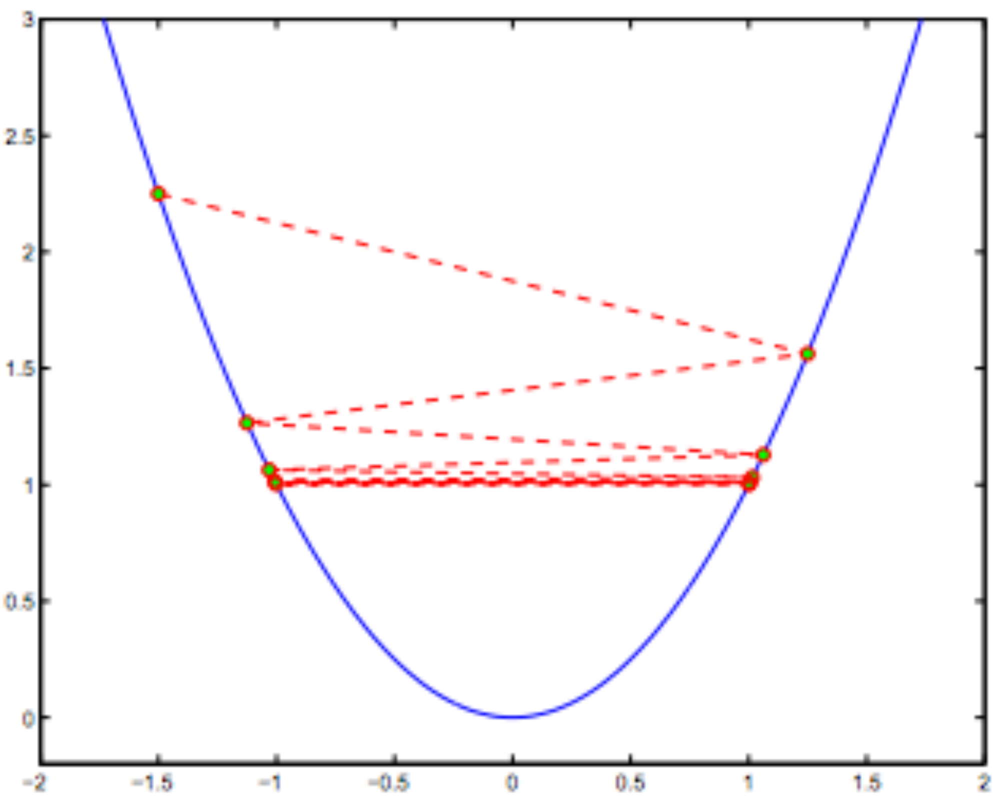
return $x^{(k+1)}$

Example: $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2), \;\; \gamma > 0$

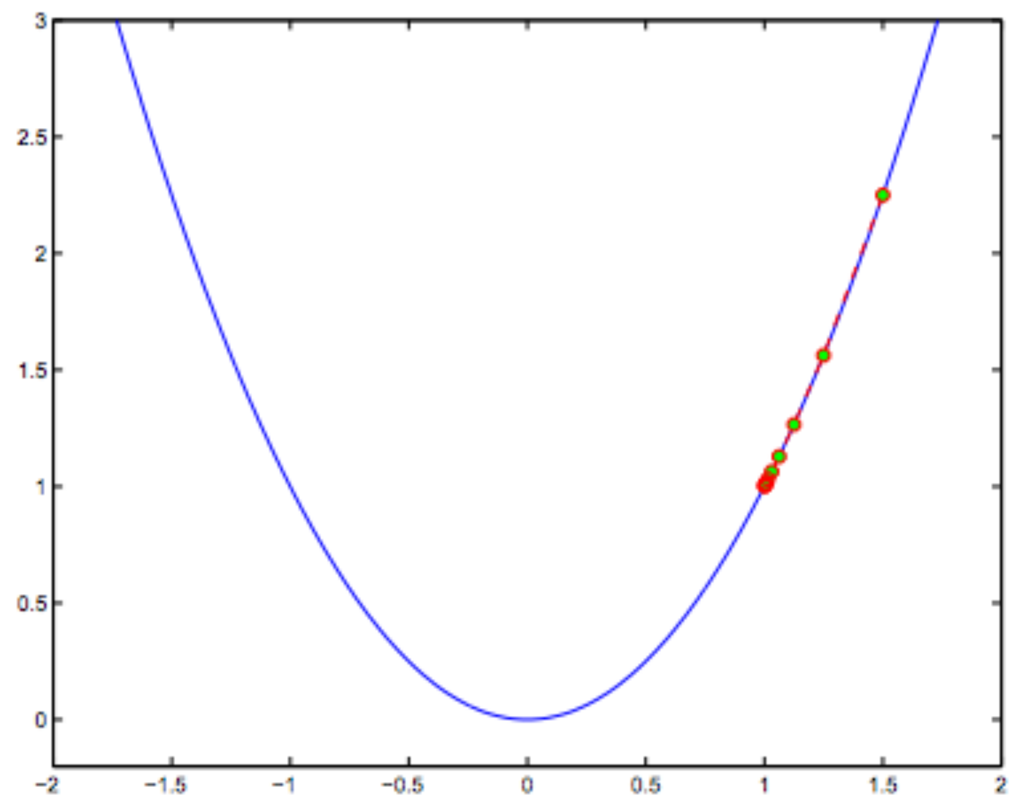


Importance of Step Size

- Challenge is to find a good step size to avoid step size that is too long or too short



too long => divergence



too short => slow convergence

Step Size Selection

- Exact Line Search: Pick step size to minimize the function

$$\eta^{(k)} = \arg\min_{\eta} f(x - \eta \nabla f(x))$$

Too expensive to be practical

- Backtracking (Armijo) Line Search: Iteratively shrink the step size until a decrease in objective is observed

Algorithm 1: Backtracking Line Search

Let $\alpha \in (0, \frac{1}{2}), \beta \in (0, 1)$
while $f(x - \eta \nabla f(x)) > f(x) - \alpha \eta ||\nabla f(x)||^2$ **do**
    $\eta = \beta \eta$
end

Example: Linear Regression

- Optimization problem:

$$\min_w ||Xw - y||_2$$

- Closed form solution:

$$w^* = (X^\top X)^{-1} X^\top y$$

- Gradient update:

$$w^+ = w - \frac{1}{m} \sum_i (x_i^\top w - y_i) x_i$$

Newton's Method

- Assumes function is locally quadratic:

$$f(x + \Delta x) \approx f(x) + \nabla f(x)^\top \Delta x + \frac{1}{2} \Delta x^\top \nabla^2 f(x) \Delta x$$

- Choose step direction:

$$\Delta x = -[\nabla^2 f(x)]^{-1} \nabla f(x)$$

- Method is often faster than gradient descent
- Hessian maybe hard and expensive to compute

Constrained Optimization Algorithms

$$\min_{x} \; f_0(x)$$
$$\text{s.t } f_k(x) \leq 0, \; k = 1, \cdots, K$$

Penalty Function Method

- Convert to one or more unconstrained optimization problems
- Penalty functions techniques:
  - Append penalty for violating constraints (exterior penalty methods)
  - Append penalty as you approach infeasibility (interior point methods)

Exterior Penalty Methods

- Linear penalty (inequality):

$$\phi_k(x) = \max(0, f_k(x))$$

- Quadratic penalty (inequality):

$$\phi_k(x) = [\max(0, f_k(x))]^2$$

- Absolute penalty (equality):

$$\phi_j(x) = \sum_j |h_j(x)|^q, \;\; q \geq 1$$

Exterior Penalty Properties

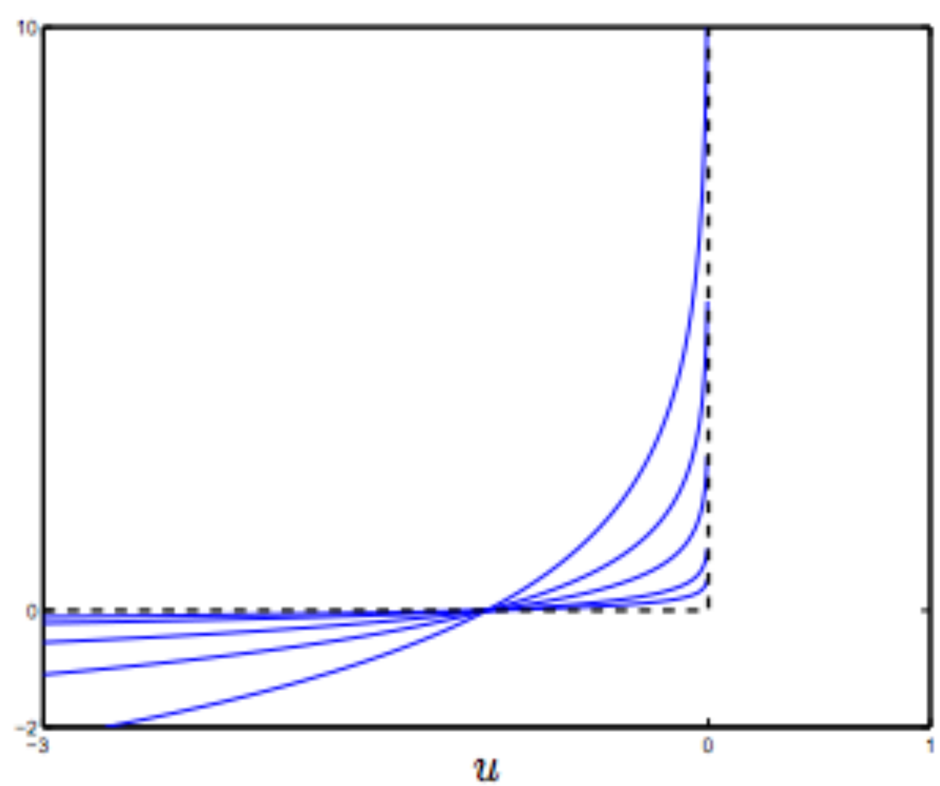
- If all constraints are satisfied, the extra terms are zero
- When penalty parameter is small, easy to minimize but yields large constraint violations
- When penalty parameter is large, constraints are nearly satisfied but optimization problem is numerically ill-conditioned
- May yield slightly infeasible solutions

Logarithmic Barrier Method

- Approximate an indicator function with log function:

$$\min_x \; f_0(x) - t\sum_k \log(-f_k(x))$$

- Solve sequence of smooth unconstrained problems with t increasing



As t gets smaller, approaches indicator function

Logarithmic Barrier Method Properties

- Convergence under mild conditions
- Barrier function is rather ill-behaved for small t
- Strictly feasible initial guess is required
- Never obtain exact solutions with active constraints but remains in the feasible space
- Only works for inequality constraints

Projected Gradient Descent

- Constrained optimization subject to convex set

$$\min \; f(x)$$
$$\text{s.t. } x \in C$$

- Projected gradient descent step:

$$x^{(k+1)} = P_C(x^{(k)} - \eta^{(k)} \nabla f(x^{(k)}))$$

- Projection onto a set c is:

$$P_C(x) = \arg\min_{v \in C} ||x - v||$$

Some Resources for Convex Optimization

- Boyd & Landenberghe's Book on Convex Optimization
https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

- Stephen Boyd's Class at Stanford
http://stanford.edu/class/ee364a/

- Vandenberghe's Class at UCLA
http://www.seas.ucla.edu/~vandenbe/ee236b/ee236b.html

- Ben-Tai & Nemirovski Lectures on Modern Convex Optimization
http://epubs.siam.org/doi/book/10.1137/1.9780898718829