Constrained Optimization

- Optimize an objective function
- Subject to conditions expressed as equalities or inequalities

\[
\begin{align*}
\text{minimize} & \quad f(x), & \text{objective} \\
\text{w.r.t} & \quad x, & \text{variables} \\
\text{subject to} & \quad a \leq x \leq b, & \text{bound constraints} \\
& \quad c_i(x) \leq u_i, & \text{inequality constraints} \\
& \quad d_j(x) = v_j. & \text{equality constraints}
\end{align*}
\]
Example: Univariate Constrained Optimization

Objective: $f(x) = (x - 2)^2 + 1$
Example: Univariate Constrained Optimization

Objective: \( f(x) = (x - 2)^2 + 1 \)

Constraint: \( c(x) = \sqrt{x} - 1 \leq 0 \)

Bound: \( x \geq 0 \)
Example: Univariate Constrained Optimization

Objective: $f(x) = (x - 2)^2 + 1$

Constraint: $c(x) \leq \sqrt{x} - 1$

Bound: $x \geq 0$

Feasible Region: $0 \leq x \leq 1$

Infeasible Region: $x > 1$
Solution Methods

- Basic idea: convert to one or more unconstrained optimization problems

- Penalty function methods
  - Append a penalty for violating constraints (exterior penalty methods)
  - Append a penalty as you approach infeasibility (interior point methods)

- Method of Lagrange multipliers
Penalty-Function Methods

1. Initialize penalty parameter
2. Initialize solution guess
3. Minimize penalized objective starting from guess
4. Update guess with the computed optimum
5. Go to 3., repeat
Linear Exterior Penalty Function

\[ \phi_i(x) = \max(0, c_i(x) - u_i) \]
Quadratic Exterior Penalty Function

\[ \phi_i(x) = \left( \max(0, c_i(x) - u_i) \right)^2 \]

Penalized objectives

Penalty function
Interior-Point Methods

Barrier function

\[ \pi(x, \mu) = f(x) - \mu \log(u_i - c_i(x)) \]

Penalized objectives

Barrier function

Barrier parameter
Summary of Penalty Function Methods

- Quadratic penalty functions always yield slightly infeasible solutions
- Linear penalty functions yield non-differentiable penalized objectives
- Interior point methods never obtain exact solutions with active constraints
- Optimization performance tightly coupled to heuristics: choice of penalty parameters and update scheme for increasing them.
- Ill-conditioned Hessians resulting from large penalty parameters may cause numerical problems
Lagrange Multipliers: Introduction

• Powerful method with deep interpretations and implications

• Append each constraint function to the objective, multiplied by a scalar for that constraint called a Lagrange multiplier. This is the Lagrangian function

\[ \mathcal{L}(x, \lambda) = f(x) + \sum_{i} \lambda_i(c_i(x) - u_i) \]

• Solution to the original constrained problem is deduced by solving for both an optimal \( x \) and an optimal set of Lagrange multipliers

• The original variables \( x \) are called primal variables, whereas the Lagrange multipliers are called dual variables

• Duality theory is both useful and beautiful, but beyond the scope of this class
Lagrange Multipliers: Motivation

There is an optimal $\lambda$ for which we obtain the constrained solution in $x$ by minimizing the Lagrangian for that $\lambda$

$$\mathcal{L}(x, \lambda) = f(x) + \lambda c(x)$$

Lagrangians for various values of the Lagrange multiplier
On to Multivariate Problems

- What changes from univariate (1-D) to multivariate (n-D) problems?
- The little pictures you’ve been seeing get very complicated very quickly
- All the concepts still work, but need more careful treatment
- Absolutely essential are concepts of level curves and gradients
Level Curves and Gradients

- Consider a function \( f : x \to y \)

- The level curves of \( f \) are curves in \( x \)-space

\[
S_k = \{ x \mid f(x) = k \}
\]

- The gradient of \( f \) w.r.t \( x \) is a vector in \( x \)-space

\[
\nabla_x(f) = \left[ \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right]
\]

- Important fact from high-school calculus:

*The gradient of a function is perpendicular to its level curves*
Level Curves and Gradients

Contour line

Gradient direction

Tangent plane
Gradients and First Order Changes

- Taylor series expansions: watch the dimensions of vectors and matrices!
  
  \[ f(x_0 + \Delta x) = f(x_0) + \left[ \nabla_x(f)|_{x_0} \right]^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2_x(f)|_{x_0} \Delta x + O(\Delta x^3) \]

- The gradient defines local tangent plane and its ‘slope.’ We can deduce that
  
  - To first order, if a change in \( x \) has a component along the gradient, \( f \) will change
  
  - To first order, there is no change in \( f \) when moving perpendicular to the gradient
  
  - By definition, there is no change in \( f \) when moving along its level curve
  
  - Hence the level curve is perpendicular to the gradient

- The Hessian defines local curvature of \( f \)
Multivariate Equality-Constrained Optimization

minimize \( x_1 + x_2, \)

subject to \( x_1^2 + x_2^2 = 4. \)

Lowest point on this surface

Provided the value on this surface is 0
Conditions for a Constrained Extremum

- Choose $x$ anywhere on the circle, i.e., at a feasible point
- Any feasible small step in $x$ must be perpendicular to the constraint gradient
- As long this step is not perpendicular to the objective gradient, we will get a change in $f$, and thereby, we at most have to reverse direction to reduce $f$
- The only way $f$ can stop changing is when the step is perpendicular to both the objective and constraint gradients
- This means that the objective gradient and constraint gradient are parallel

\[ \nabla_x(f)|_{x^*} = \lambda \nabla_x(c)|_{x^*} \]

- We have just found a constrained local extremum of $f$
Question: what if some components of the constraint gradient are zero?