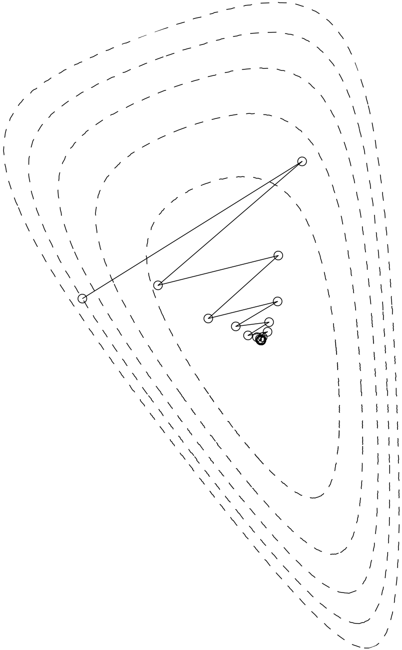


Optimization Algorithms

Non-Linear Mathematical Programs & KKT

Marc Toussaint
Technical University of Berlin
Winter 2024/25

Problem Formulation

- General **Non-linear Mathematical Program (NLP)** (constrained optimization problem):
  Let $x \in \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R}^n \to \mathbb{R}^m$, $h: \mathbb{R}^n \to \mathbb{R}^l$ find

$$\min_x \; f(x) \;\text{ s.t. }\; g(x) \le 0,\; h(x) = 0$$

  – We typically assume f, g, h to be differentiable or smooth.
  – We can typically query $f, \nabla f, g, \nabla g, h, \nabla h$, optionally also $\nabla^2 f$.

- The lecture sometimes only mentions inequality constraints g, equality constraints are analogous/easier

General approaches

- Roughly, try to somehow transform the constraint problem to
  - a series of unconstraint problems (log-barrier, AugLag, etc...)
  - a single but larger unconstraint problem (primal-dual)
  - a (series of) other constraint problems, hopefully simpler (dual, convex, SQP)

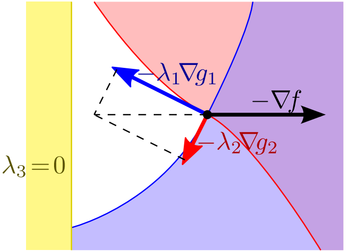
Outline

- KKT conditions of optimality
- Core methods: **log barrier**, squared penalties, **Aug. Lagrangian**
- Introduce **Lagrangian** – revisit KKT, log barrier, dual problem, primal-dual
- Further topics: Phase I, bound constraints, trust region, distributed optimization, simplex algorithm

Solving by sketching

- Sketch the following problems and identify the solution:
  - 1-dimensional: $\min x$ s.t. $\sin(x) = 0,\ x^2/4 - 1 \le 0,$
  - 2-dimensional: $\min x_1$ s.t. $x^2 + y^2 - 1 \le 0$

- *At the optimum there must be a balance between the cost gradient $-\nabla f(x)$ pulling, and the gradient of the active constraints $-\nabla g_i(x)$ pushing back*



[Our convention: "costs $f(x)$ *pull*, constraints $g(x), h(x)$ *push* back"]

Karush-Kuhn-Tucker conditions

- *At the optimum there must be a balance between the cost gradient $-\nabla f(x)$ pulling, and the gradient of the active constraints $-\nabla g_i(x)$ pushing back*

Karush-Kuhn-Tucker conditions

- *At the optimum there must be a balance between the cost gradient* $-\nabla f(x)$ *pulling, and the gradient of the active constraints* $-\nabla g_i(x)$ *pushing back*

- **Theorem (Karush-Kuhn-Tucker conditions):** For any NLP,

$$x \text{ optimal} \quad \Rightarrow \quad \exists \lambda \in \mathbb{R}^m, \kappa \in \mathbb{R}^l \text{ s.t.}$$

$$\nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{j=1}^l \kappa_j \nabla h_j(x) = 0 \qquad \text{(stationarity)}$$

$$h(x) = 0 \;, \quad g(x) \le 0 \qquad \text{(primal feasibility)}$$

$$\lambda \ge 0 \qquad \text{(dual feasibility)}$$

$$\forall_i : \; \lambda_i g_i(x) = 0 \qquad \text{(complementarity)}$$

[stationarity in compact notation: $\nabla f(x) + \lambda^\top \partial_x g(x) + \kappa^\top \partial_x h(x) = 0$]

Karush-Kuhn-Tucker conditions

- **Stationarity (1st KKT)**: “Force balance” of the cost pulling and the active constraints pushing back
  - Existence of dual parameters λ, κ is equivalent to

$$\nabla f(x) \in \text{span}\{\nabla g_{1..m}, \nabla h_{1..l}\}$$

  - The values of λ and κ $\leftrightarrow$ how strongly the constraints push

Karush-Kuhn-Tucker conditions

- **Stationarity (1st KKT)**: "Force balance" of the cost pulling and the active constraints pushing back
  - Existence of dual parameters λ, κ is equivalent to

$$\nabla f(x) \in \text{span}\{\nabla g_{1..m}, \nabla h_{1..l}\}$$

  - The values of λ and κ $\leftrightarrow$ how strongly the constraints push

- **Complementarity (4th KKT)**: "Logic of constraint activity"
  - An inequality can only push at the boundary, where $g_i = 0$
  - The formulation $\lambda_i g_i = 0$ very elegantly describes this logic
  - The combinatorics of which constraint is active ($O(2^m)$) is a core source of difficulty of constrained optimization
  - If you would apriori know which constraints are active $\rightarrow$ inequalities become equalities $\rightarrow$ easier