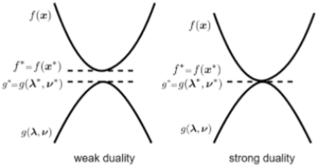
Optimization in Machine Learning

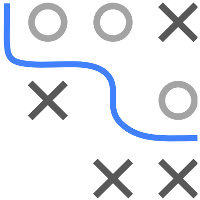
Nonlinear programs and Lagrangian



Learning goals

- Lagrangian for general constrained optimization
- Geometric intuition for Lagrangian duality
- Properties and examples

NONLINEAR CONSTRAINED OPTIMIZATION



Previous lecture: **Linear programs**

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^d} \quad & f(\mathbf{x}) := \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{G}\mathbf{x} = \mathbf{h} \end{aligned}$$

Related to its (Lagrange) dual formulation by the *Lagrangian*

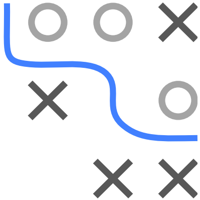
$$\mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{c}^\top \mathbf{x} + \boldsymbol{\alpha}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\beta}^\top(\mathbf{G}\mathbf{x} - \mathbf{h}).$$

Weak duality: For $\boldsymbol{\alpha} \geq 0$ and $\boldsymbol{\beta}$:

$$f(\mathbf{x}^*) \geq \min_{\mathbf{x}\in\mathcal{S}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \geq \min_{\mathbf{x}\in\mathbb{R}^d} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) =: g(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

Recall: Implicit domain constraint in *Lagrange dual function* $g(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

CONSTRAINED PROBLEMS: THE DIRECT WAY



Simple constrained problems can be solved in a direct way.

Example 1:

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & 2 - x^2 \\ \text{s.t.} \quad & x - 1 = 0 \end{aligned}$$

Solution: Resolve the constraint by

$$\begin{aligned} x - 1 &= 0 \\ x &= 1 \end{aligned}$$

and insert it into the objective:

$$x^* = 1, \quad f(x^*) = 1$$

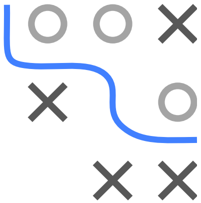
A CLASSIC EXAMPLE: "MILKMAID PROBLEM"

Question 1: Is there a general recipe for solving general constrained nonlinear optimization problems?
Question 2: Can we understand this recipe geometrically?
Question 3: How does this relate to the Lagrange function approach?

For this purpose, we consider the classical “milkmaid problem”; the following example is taken from *Steuard Jensen, An Introduction to Lagrange Multipliers* (but the example works of course equally well with a “milk man”).

- Assume a milk maid is sent to the field to get the day’s milk
- The milkmaid wants to finish her job as quickly as possible
- However, she has to clean her bucket first at the nearby river.

LAGRANGE FUNCTION



General: Solve problem with single equality constraint by:

$$\begin{aligned} \nabla f(\mathbf{x}) &= \beta \nabla h(\mathbf{x}) \\ h(\mathbf{x}) &= 0 \end{aligned}$$

- **First line**: Parallel gradients | **Second line:** Constraint

Observe: Above system is equivalent to

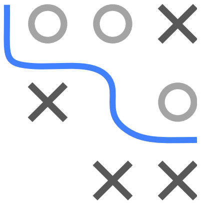
$$\nabla \mathcal{L}(\mathbf{x}, \beta) = \mathbf{0}$$

for **Lagrange function** (or **Lagrangian**) $\mathcal{L}(\mathbf{x}, \beta) := f(\mathbf{x}) + \beta h(\mathbf{x})$

Indeed:

$$\begin{pmatrix} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \beta) \\ \nabla_{\beta} \mathcal{L}(\mathbf{x}, \beta) \end{pmatrix} = \begin{pmatrix} \nabla f(\mathbf{x}) + \beta \nabla h(\mathbf{x}) \\ h(\mathbf{x}) \end{pmatrix}$$

LAGRANGE FUNCTION AND PRIMAL PROBLEM



General constrained optimization problems:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, k \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, \ell \end{aligned}$$

Extend Lagrangian ($\alpha_i \geq 0$, β_i Lagrange multipliers):

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) := f(\mathbf{x}) + \sum_{i=1}^{k} \alpha_i g_i(\mathbf{x}) + \sum_{j=1}^{\ell} \beta_j h_j(\mathbf{x})$$

Equivalent primal problem:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

Question: Why?

EXAMPLE: LAGRANGE FUNCTION FOR QP'S

We consider quadratic programming

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} \\ \text{s.t.} \quad & h(\mathbf{x}) := \mathbf{C}\mathbf{x} - \mathbf{d} = \mathbf{0} \end{aligned}$$

with $\mathbf{Q} \in \mathbb{R}^{d \times d}$ symmetric, $\mathbf{C} \in \mathbb{R}^{\ell \times d}$, and $\mathbf{d} \in \mathbb{R}^{\ell}$.

Lagrange function: $\mathcal{L}(\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \boldsymbol{\beta}^\top(\mathbf{C}\mathbf{x} - \mathbf{d})$

Solve

$$\begin{aligned} \nabla \mathcal{L}(\mathbf{x}, \boldsymbol{\beta}) = \begin{pmatrix} \partial \mathcal{L}/\partial \mathbf{x} \\ \partial \mathcal{L}/\partial \boldsymbol{\beta} \end{pmatrix} &= \begin{pmatrix} \mathbf{Q}\mathbf{x} + \mathbf{C}^\top \boldsymbol{\beta} \\ \mathbf{C}\mathbf{x} - \mathbf{d} \end{pmatrix} = \mathbf{0} \\ \Leftrightarrow \qquad \begin{pmatrix} \mathbf{Q} & \mathbf{C}^\top \\ \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\beta} \end{pmatrix} &= \begin{pmatrix} \mathbf{0} \\ \mathbf{d} \end{pmatrix} \end{aligned}$$

Observe: Solve QP by solving a linear system

LAGRANGE DUALITY

Dual problem:

$$\max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

Define **Lagrange dual function** $g(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$

Important characteristics of the dual problem:

- **Convexity** (pointwise minimum of *affine* functions)
    - Gives methods based on dual solutions
    - Might be computationally inefficient (expensive minimizations)
- **Weak duality:**

$$f(\mathbf{x}^*) \geq g(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$$

- **Strong duality** if primal problem satisfies *Slater's condition*[(1)]:

$$f(\mathbf{x}^*) = g(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$$

[(1)] **Slater's condition**: Primal problem convex and "strictly feasible" ($\exists \mathbf{x} \forall i : g_i(\mathbf{x}) < 0$).