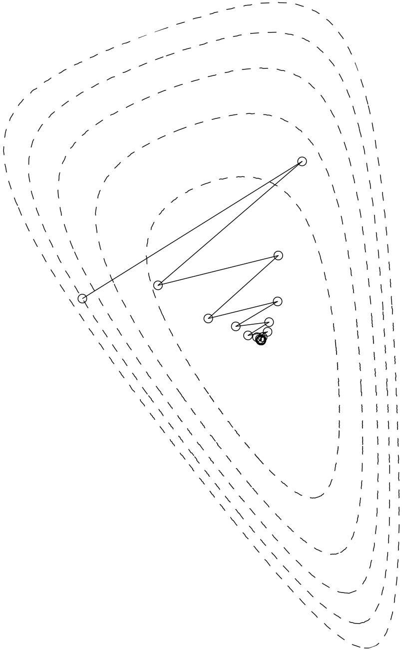

Optimization Algorithms

The Lagrangian

Definition, Relation to KKT conditions, saddle point view, dual problem, min-max max-min duality, modified KKT & log barriers

Marc Toussaint
Technical University of Berlin
Winter 2024/25

The Lagrangian

- Given a constraint problem

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

  we define the **Lagrangian** as

$$\begin{aligned} L(x,\kappa,\lambda) &= f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^l \kappa_i h_i(x) \\ &= f(x) + \lambda^\top g(x) + \kappa^\top h(x) \end{aligned}$$

- The $\lambda_i \ge 0$ and $\kappa_i \in \mathbb{R}$ are called **dual variables** or *Lagrange multipliers*

What's the point of this definition?

- The Lagrangian relates strongly to the KKT conditions of optimality!
- The Lagrangian is useful to compute optima analytically, on paper
- Optima are necessarily at saddle points of the Lagrangian
- The Lagrangian implies a dual problem, which is sometimes easier to solve than the primal

Relations between Lagrangian and KKT

- $\nabla_x L = 0$ implies the **1st KKT** condition

$$0 = \nabla_x L = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^l \kappa_i \nabla h_i(x)$$

Relations between Lagrangian and KKT

- $\nabla_x L = 0$ implies the **1st KKT** condition

$$0 = \nabla_x L = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^l \kappa_i \nabla h_i(x)$$

- $\nabla_\kappa L = 0$, implies primal feasibility ($h = 0$, **2nd KKT**) w.r.t. the equalities

Relations between Lagrangian and KKT

- $\nabla_x L = 0$ implies the **1st KKT** condition

$$0 = \nabla_x L = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{i=1}^l \kappa_i \nabla h_i(x)$$

- $\nabla_\kappa L = 0$, implies primal feasibility ($h = 0$, **2nd KKT**) w.r.t. the equalities
- $\max_{\lambda \geq 0} L$ is related to the remaining **2nd and 4th KKT** conditions:

$$\max_{\lambda \geq 0} L(x, \lambda) = F_\infty(x) \overset{\text{def}}{=} \begin{cases} f(x) & \text{if } g(x) \leq 0 \\ \infty & \text{otherwise} \end{cases} \tag{1}$$

$$\lambda = \underset{\lambda \geq 0}{\text{argmax}}\, L(x, \lambda) \quad \Rightarrow \quad \begin{cases} \lambda_i = 0 & \text{if } g_i(x) < 0 \\ 0 = \nabla_{\lambda_i} L(x, \lambda) = g_i(x) & \text{otherwise} \end{cases} \tag{2}$$

This implies either $(\lambda_i = 0 \wedge g_i(x) < 0)$ or $g_i(x) = 0$, which is equivalent to the *complementarity* and *primal feasibility* for inequalities.

Relations between Lagrangian and KKT

- We learnt that
  - $\min_x L(x, \lambda, \kappa)$ reproduces 1st KKT
  - $\max_{\lambda \geq 0, \kappa} L(x, \lambda, \kappa)$ reproduces remaining KKT
- KKT conditions are related to minimize w.r.t. x, and maximize w.r.t. λ... (more on this later)

Relations between Lagrangian and KKT

- We learnt that
  - $\min_x L(x, \lambda, \kappa)$ reproduces 1st KKT
  - $\max_{\lambda \geq 0, \kappa} L(x, \lambda, \kappa)$ reproduces remaining KKT
- KKT conditions are related to minimize w.r.t. x, and maximize w.r.t. λ... (more on this later)
- How can we use this?
  - The KKT conditions state that, at an optimum, there exist some λ, κ. This existance statement is not directly helpful to actually find them.
  - In contrast, the Lagrangian tells us how the dual parameters can be found: by maximizing w.r.t. them.

Solving constraint problems on paper

- For $x \in \mathbb{R}^2$, what is

$$\min_x x^2 \text{ s.t. } x_1 + x_2 = 1$$

Solving constraint problems on paper

- For $x \in \mathbb{R}^2$, what is

$$\min_x x^2 \text{ s.t. } x_1 + x_2 = 1$$

- Solution:

$$L(x, \kappa) = x^2 + \kappa(x_1 + x_2 - 1)$$

$$0 = \nabla_x L(x, \kappa) = 2x + \kappa \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \Rightarrow \quad x_1 = x_2 = -\kappa/2$$

$$0 = \nabla_\kappa L(x, \kappa) = x_1 + x_2 - 1 = -\kappa/2 - \kappa/2 - 1 \quad \Rightarrow \quad \kappa = -1$$

$$\Rightarrow x_1 = x_2 = 1/2$$

  – κ is also called *Lagrange multiplier*, I prefer *dual variables*
  – When applying this to inequalities, you have to consider *both cases* (inequality active, and inequality inactive) and check if the inactive solution is feasible ($g \le 0$) or the active solution dual-feasible ($\lambda \ge 0$). (For m inequality constraints, you run into 2^m combinatorics.)

Saddle Points, Primal & Dual Problems

[For simplicity, consider inequalities only.]

- $\min_x L(x, \lambda)$ reproduces 1st KKT; $\max_{\lambda \geq 0} L(x, \lambda)$ reproduces remaining KKT

Saddle Points, Primal & Dual Problems

[For simplicity, consider inequalities only.]

- $\min_x L(x, \lambda)$ reproduces 1st KKT; $\max_{\lambda \geq 0} L(x, \lambda)$ reproduces remaining KKT

- This motivates defining the **Primal and dual problem** (details later):

$$\min_x \underbrace{\max_{\lambda \geq 0} L(x, \lambda)}_{F_\infty(x) \quad (\infty\text{-barrier function})} \qquad \textbf{(primal problem)}$$

$$\max_{\lambda \geq 0} \underbrace{\min_x L(x, \lambda)}_{l(\lambda) \quad \textbf{(dual function)}} \qquad \textbf{(dual problem)}$$

  – Convince yourself, that the first problem is the original problem $\min_x f(x)$ s.t. $g(x) \leq 0$
  – Find a tabular function $L(x, \lambda)$ (for discrete $x, \lambda \in \{1, 2\}$) where $\min_x \max_\lambda L \neq \max_\lambda \min_x L$

The Lagrange dual problem

$$\min_x \underbrace{\max_{\lambda \ge 0} L(x,\lambda)}_{F_\infty(x)} \qquad \textbf{(primal problem)}$$

$$\max_{\lambda \ge 0} \underbrace{\min_x L(x,\lambda)}_{l(\lambda)} \qquad \textbf{(dual problem)}$$

- We defined the **dual function** as

$$l(\lambda) = \min_x L(x,\lambda)$$

- **Theorem:** The dual problem is convex (objective=concave, constraints=convex), even if the primal is non-convex!
  - $L(x,\lambda)$ is linear in λ
  - $l(\lambda) = \min_x L(x,\lambda)$ is a point-wise minimization $\Rightarrow l(x)$ concave

The Lagrange dual problem

- Sometimes, $l(\lambda) = \min_x L(x, \lambda)$ can be derived analytically. We could swap a non-convex primal problem for a convex dual problem. However, in general $\min_x \max_y f(x, y) \neq \max_y \min_x f(x, y)$.

The Lagrange dual problem

- Sometimes, $l(\lambda) = \min_x L(x,\lambda)$ can be derived analytically. We could swap a non-convex primal problem for a convex dual problem. However, in general $\min_x \max_y f(x,y) \neq \max_y \min_x f(x,y)$.

- The dual function is always a **lower bound** (for $\lambda_i \geq 0$):

$$\lambda_i \geq 0 \quad \Rightarrow \quad L(x,\lambda) \leq F_\infty(x)$$

$$l(\lambda) = \min_x L(x,\lambda) \;\leq\; \min_x F_\infty(x) = p^* \overset{\text{def}}{=} \left[\min_x f(x) \;\text{ s.t. }\; g(x) \leq 0\right]$$

$$\max_{\lambda \geq 0} \min_x L(x,\lambda) \;\leq\; \min_x \max_{\lambda \geq 0} L(x,\lambda) = p^*$$

$$l(\lambda) \leq p^*$$

The Lagrange dual problem

- We say **strong duality** holds iff

$$\max_{\lambda \geq 0} \min_x L(x, \lambda) = \min_x \max_{\lambda \geq 0} L(x, \lambda)$$

- **Theorem:** If the primal is convex, and there exist an interior point

$$\exists_x : \; \forall_i : \; g_i(x) < 0$$

(which is called **Slater condition**), then we have *strong duality* (Boyd, Sec 5.3.2)

Log barrier method revisited

Log barrier method revisited

- Recall, the inner iterations minimize $\min_x f(x) - \mu \sum_i \log(-g_i(x))$:

$$\nabla f(x) + \textstyle\sum_i \lambda_i \nabla g_i(x) = 0 \;, \quad \text{with } \lambda_i g_i(x) = -\mu$$

- With the definition $\lambda_i = -\mu / g_i(x^*)$ and $x^*(\mu) = \mathrm{argmin}_x\, B(x,\mu)$, we have

$$\nabla B(x,\mu) = \nabla f(x) + \textstyle\sum_{i=1}^m \lambda_i \nabla g_i(x) = \nabla L(x,\lambda)$$
$$x^*(\mu) = \operatorname*{argmin}_x L(x,\lambda) \;, \quad \text{with } \lambda_i g_i(x) = -\mu$$

- We also have (with m the count of inequalities)

$$l(\lambda) = \min_x L(x,\lambda) = f(x^*) + \textstyle\sum_{i=1}^m \lambda_i g_i(x^*) = f(x^*) - m\mu$$

- Further, as the dual function is a lower bound, $l(\lambda) \le p^*$, we have

$$f(x^*) - p^* \le m\mu$$

μ is an upper bound on the suboptimality of the centering point x^*

Log barrier method – Conclusions

- The μ, which we introduced as factor for the log barrier, has "deep semantics":

- μ defines a **relaxation of the 4th KKT** complementarity condition
- the log barrier gradients $\lambda_i = -\mu / g_i(x^*)$ have the semantics of dual variables
- $x^*(\mu)$ solves the relaxed KKT
- $f(x^*(\mu)) = l(\lambda) + m\mu$ gives the dual function value for λ
- μ defines an upper bound on the **sub-optimality** of each x^*: $\quad f(x^*) - p^* \le m\mu$

Comments

- We first learnt about three basic methods to tackle constrained optimization by repeated unconstrained optimization:
  - Log barrier method
  - Squared penalty method (approximate only)
  - Augmented Lagrangian method

- We understood KKT, Lagrangian, dual problem, saddle point view, duality gap, relation to μ