OM-S20-26: Comments on Lagrangian Duality (optional)

C. V. Jawahar

IIIT Hyderabad

http://preon.iiit.ac.in/om_quiz

24 Apr 2020

Announcements

- Course getting over. Corner and special cases are in focus.
- HW and Assignment marks partly on moodle. Others will be there by early next week. Talk to Kritika and Aditya on this.
- Class Reviews: (i) You had seen the previous scores. (ii) two optional to replace unexpected situations (including today). Final scores should be there on moodle by weekend. Nearly closed. PoC: Piyush
- ILM: (i) ILM2 discussins will be closed today. ILM3 discussions will be open till 29th. (ii) Scores will start coming by Monday and close by May 1st. PoC: Jawahar
- Special permission from Acad. office. (shall get back to them by middle of next week with decisions/plans. If you do not get any email from us, please write back.) PoC: Jawahar

KKT: Recap

$$\max f(x) \quad \text{s.t}$$
$$g_j(x) \leq 0 \;\; j = 1, \ldots, m$$
$$h_j(x) = 0 \;\; j = 1, \ldots, l$$

Lagrangian:

$$L(x, \lambda, \mu) = f(x) - \sum_{j=1}^{m} \lambda_j g_j(x) - \sum_{j=1}^{l} \mu_j h_j(x)$$

Then at x^*:

$$\nabla f = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) + \sum_{j=1}^{l} \mu_j \nabla h_j(x^*)$$

$$g_j(x^*) \leq 0 \text{ and } h_j(x^*) = 0$$
$$\lambda_j g_j(x^*) = 0 \;\; j = 1, \ldots, m$$
$$\lambda_j \geq 0 \text{ and } \mu_j \in R$$

Lagrangian Duality

$$\min_x f(x)$$
$$s.t. \quad g_j(x) \le 0 \quad , j = 1 \ldots n$$
$$h_i(x) = 0 \quad , j = 1 \ldots m$$

is equivalent to the following optimization problem,

Primal: $$\min_x \max_{\mu \ge 0, \lambda} \underbrace{f(x) + \sum_j \mu_j g_j(x) + \sum_i \lambda_i h_i(x)}_{L(x, \mu \ge 0, \lambda)}$$

Because,

$$\max_{\mu \ge 0, \lambda} L(x, \mu, \lambda) = \begin{cases} f(x) & when \quad x \quad is \quad feasible \\ \infty & otherwise \end{cases}$$

Lagrangian Duality

$$\min_x f(x)$$
$$s.t. \quad g_j(x) \le 0 \quad , j = 1 \ldots n$$
$$h_i(x) = 0 \quad , j = 1 \ldots m$$

$$|||$$

Primal: $$\min_x \max_{\mu \ge 0, \lambda} \underbrace{f(x) + \sum_j \mu_j g_j(x) + \sum_i \lambda_i h_i(x)}_{L(x, \mu \ge 0, \lambda)}$$

Dual: $$\max_{\mu \ge 0, \lambda} \min_x f(x) + \sum_j \mu_j g_j(x) + \sum_i \lambda_i h_i(x)$$

Dual *Lower bound of* Primal : $\max_{\mu \ge 0, \lambda} \min_x L(.) \le \min_x \max_{\mu \ge 0, \lambda} L(.)$

Lagrangian Duality

- Min-Max Theorem says

$$\max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} \phi(x, y) \leq \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \phi(x, y)$$

- In our case of Lagrangian

$$\max_{x \in \mathcal{X}} \min_{\lambda \in \Lambda} L(x, \lambda) \leq \min_{\lambda \in \Lambda} \max_{x \in \mathcal{X}} L(x, \lambda)$$

- $d^* \leq p^*$
- Weak Duality: The optima of the maximization problem is inferior or equal to the optima of the minimization problem.
- Strong Duality: In special cases (like LP), we know that it is strong duality where $\leq$ is replaced by $=$.

LD: Example 1: LP

Consider the primal problem

$$p^* = \min c^T x \text{ s. t. } Ax \leq b$$

The corresponding Lagrangian is:

$$L(x, \lambda) = c^T x + \lambda^T [Ax - b]$$

We know

$$g(\lambda) = \min_x L(x, \lambda) = \begin{cases} -b^T \lambda & \text{if } A^T \lambda + c = 0 \\ +\infty & \text{otherwise} \end{cases}$$

$$d^* = \max_\lambda -b^T \lambda : \lambda \geq 0 \quad A^T \lambda + c = 0$$

For primal LP problems with inequality constraints, dual is another LP with equality constraints.

Least Norm (Recap)

We had seen this problem:

$$\min \ \frac{1}{2}||x||_2^2 \quad \text{s.t.} \ Ax = b$$

We know the solution as

$$x^* = A^T(AA^T)^{-1}b$$

What about the optimum value

$$p^* = \frac{1}{2}x^Tx = \frac{1}{2}(A^T(AA^T)^{-1}b)^T(A^T(AA^T)^{-1}b)$$

$$= \frac{1}{2}b^T(AA^T)^{-1}b$$

LD: Example 2: Least Nrom

Consider the problem of Least norm

$$\min \ \frac{1}{2}||x||_2^2 \ \text{ s.t. } Ax = b$$

Lagrangian is

$$L(x, \lambda) = \frac{1}{2}||x||_2^2 + \lambda^T[Ax - b]$$

Lagrangian dual is

$$g(\lambda) = \min_x L(x, \lambda) = \min_x \frac{1}{2}||x||_2^2 + \lambda^T[Ax - b]$$

$$\nabla_x L(x, \lambda) = x + A^T\lambda = 0 \text{ or } x = -A^T\lambda$$

$$g(\lambda) = -\frac{1}{2}\lambda^T AA^T\lambda - b^T$$

$$d^* = \max_\lambda g(\lambda) = \frac{1}{2}b^T(AA^T)^{-1}b$$

LD: Example 3: SVM

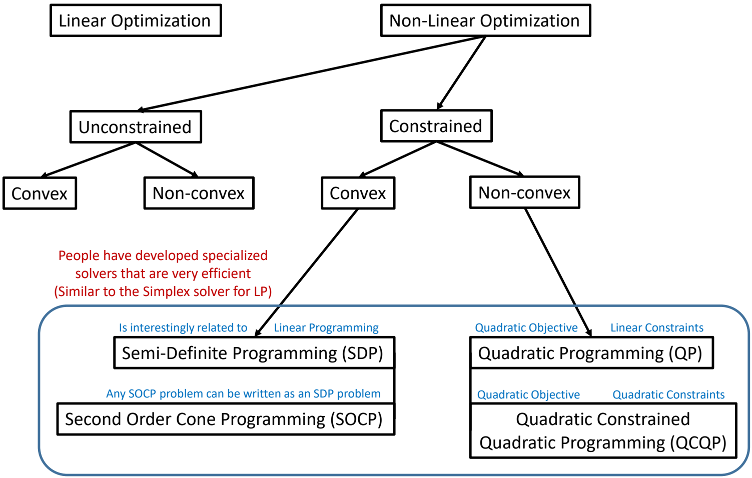
Binary SVM Learning problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i \\ \text{s.t} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i \ ; \ \xi_i \geq 0 \ \ \forall i \end{aligned}$$

Lagrangian function:

$$\begin{aligned} & \mathcal{L}(w, b, \xi, \alpha, \beta) \\ & = \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i \left[y_i (w^T x_i + b) - 1 + \xi_i \right] - \sum_{i=1}^{m} \beta_i \xi_i \end{aligned}$$

Summary: An incomplete taxonomy of problems



Summary

LP and IP Formulations	Nonlinear Optimization
LP Relaxation	Convex Optimization
Solve $Ax = b$	Solve $f(x) = 0$
LSE and Least Norm	Spectral Methods
Optimize $x^T Ax$ s.t $\|\|x\|\| = 1$	KKT and Optimality
Numerical Linear Algebra	Totally Unimodular matrices
Graphical Solns	Fixed Point Iterations
Simplex	GD and NM
LP Duality	Lagrangian Duality
Applications in Graphs	Applications in ML/DL