CH2.1: DUAL ASCENT

Monday, April 1, 2024 4:18 PM

• DUAL ASCENT

• Equality constrained optimization problem: (PRIMAL PROBLEM)

$$\begin{cases} \text{minimize } f(x) \\ \text{subject to } Ax = b \end{cases} \qquad x \in \mathbb{R}^n,\ A \in \mathbb{R}^{m \times n},\ f: \mathbb{R}^n \to \mathbb{R} \text{ is convex} \qquad (2.1)$$

• LAGRANGIAN : $L(x,y) = f(x) + y^T(Ax - b)$

• DUAL FUNCTION : $g(y) = \inf_x L(x,y) = -f^*(-A^T y) - b^T y$

y : "dual variable" or "Lagrangian multiplier".

f^* is the convex conjugate of f.

• DUAL PROBLEM : maximize $g(y)$ $\qquad y \in \mathbb{R}^n$

Assuming that strong duality holds, the optimal values of the primal and dual problems are the same.
We can recover the primal optimal point x-star from the dual optimal point y-star as:

$$x^* = \operatorname{argmin}_x L(x, y^*)$$

Assuming that there is only one minimizer of L(x,y-star). This happens when f is strictly convex.

DUAL ASCENT PROBLEM:

Solve the dual problem using gradient ascent.
Assuming that g is differentiable:

Find $\nabla g(y)$:

$x^+ = \operatorname{argmin}_x L(x,y)$, then

$\nabla g(y) = Ax^+ - b$ (residual for the equality constraint)

DUAL ASCENT METHOD : k iterations

$x^{k+1} := \operatorname{argmin}_x L(x, y^k)$ (x-minimization step)

$y^{k+1} := y^k + \alpha^k (Ax^{k+1} - b)$ (dual variable update)

$\alpha^k > 0$ is step size

- The y-variable can be interpreted as a "vector of prices".
- The y-update is then called the "price update" or "price adjustment" step.
- The algorithm is called "dual ascent" since, with an appropriate choice of alpha^k, the dual function increases in each step:

$$g(y^{k+1}) > g(y^k)$$

- DUAL SUBGRADIENT METHOD: (Non-Differentiable g)

  - The dual ascent method can also be used in some cases where g is not differentiable.
  - In this case, the residual, $Ax^{k+1} - b$ is not the gradient of g, but the negative of the subgradient of $-g$.
  - This case requires a different choice of α^k than when g is differentiable, and convergence is not monotone, as is often the case when $g(y^{k+1}) \ngtr g(y^k)$
  - In this case, the algorithm is called the "dual subgradient method".

- If α^k is chosen appropriately and several assumptions hold, then x^k converges to an optimal point and y^k converges to an optimal dual point.

↳ However, these assumptions often do not hold in many applications, so dual ascent often cannot be used.

ex) f is a non-zero affine function of any component of x, then x-update fails, since L is unbounded below in x for most y.