12. Dual proximal gradient method

- proximal gradient method applied to the dual
- examples
- alternating minimization method

Dual methods

subgradient method: slow, step size selection difficult

gradient method: requires differentiable dual cost function

- often dual cost is not differentiable, or has nontrivial domain
- dual can be smoothed by adding small strongly convex term to primal

augmented Lagrangian method:

- equivalent to gradient ascent on a smoothed dual problem
- however smoothing destroys separable structure

proximal gradient method (this lecture): dual cost split in two terms

- one term is differentiable with Lipschitz continuous gradient
- other term has an inexpensive prox-operator

Composite structure in the dual

$$\text{minimize} \quad f(x) + g(Ax) \qquad\qquad \text{maximize} \quad -f^*(-A^Tz) - g^*(z)$$

dual has the right structure for the proximal gradient method if

- prox-operator of g (or g^*) is cheap (closed form or simple algorithm)
- f is strongly convex ($f(x) - (\mu/2)x^Tx$ is convex)

  implies $f^*(-A^Tz)$ has Lipschitz continuous gradient ($L = \|A\|_2^2/\mu$):

$$\left\|A\nabla f^*(-A^Tu) - A\nabla f^*(-A^Tv)\right\|_2 \le \frac{\|A\|_2^2}{\mu}\,\|u-v\|_2$$

  because ∇f^* is Lipschitz continuous with constant $1/\mu$ (see page 11-6)

Dual proximal gradient update

$$z^+ = \operatorname{prox}_{tg^*}\left(z + tA\nabla f^*(-A^T z)\right)$$

equivalent expression in terms of f:

$$z^+ = \operatorname{prox}_{tg^*}(z + tA\hat{x}) \quad \text{where } \hat{x} = \operatorname*{argmin}_x \left(f(x) + z^T Ax\right)$$

- if f is separable, calculation of $\hat{x}$ decomposes into independent problems
- step size t constant or from backtracking line search
- can use accelerated proximal gradient methods of lecture 7

Alternating minimization interpretation

Moreau decomposition gives alternate expression for z-update

$$z^+ = z + t(A\hat{x} - \hat{y})$$

where

$$\begin{aligned}
\hat{x} &= \operatorname*{argmin}_x \, (f(x) + z^T Ax) \\
\hat{y} &= \operatorname{prox}_{t^{-1}g}(z/t + A\hat{x}) \\
&= \operatorname*{argmin}_y \, (g(y) + z^T(A\hat{x} - y) + \frac{t}{2}\|A\hat{x} - y\|_2^2)
\end{aligned}$$

in each iteration, an alternating minimization of:

- Lagrangian $f(x) + g(y) + z^T(Ax - y)$ over x
- augmented Lagrangian $f(x) + g(y) + z^T(Ax - y) + \frac{t}{2}\|Ax - y\|_2^2$ over y

Outline

- proximal gradient method applied to the dual
- **examples**
- alternating minimization method

Regularized norm approximation

$$\text{minimize} \quad f(x) + \|Ax - b\| \qquad \text{(with } f \text{ strongly convex)}$$

a special case of page 12-3 with $g(y) = \|y - b\|$

$$g^*(z) = \begin{cases} b^T z & \|z\|_* \leq 1 \\ +\infty & \text{otherwise} \end{cases} \qquad \mathrm{prox}_{tg*}(z) = P_C(z - tb)$$

C is unit norm ball for dual norm $\|\cdot\|_*$

dual gradient projection update

$$\begin{aligned} \hat{x} &= \operatorname*{argmin}_x \left(f(x) + z^T Ax\right) \\ z^+ &= P_C(z + t(A\hat{x} - b)) \end{aligned}$$

Example

$$\text{minimize} \quad f(x) + \sum_{i=1}^{p} \|B_i x\|_2 \qquad \text{(with } f \text{ strongly convex)}$$

a special case of page 12-3 with $g(y_1, \ldots, y_p) = \sum_{i=1}^{p} \|y_i\|_2$ and

$$A = \begin{bmatrix} B_1^T & B_2^T & \cdots & B_p^T \end{bmatrix}^T$$

dual gradient projection update

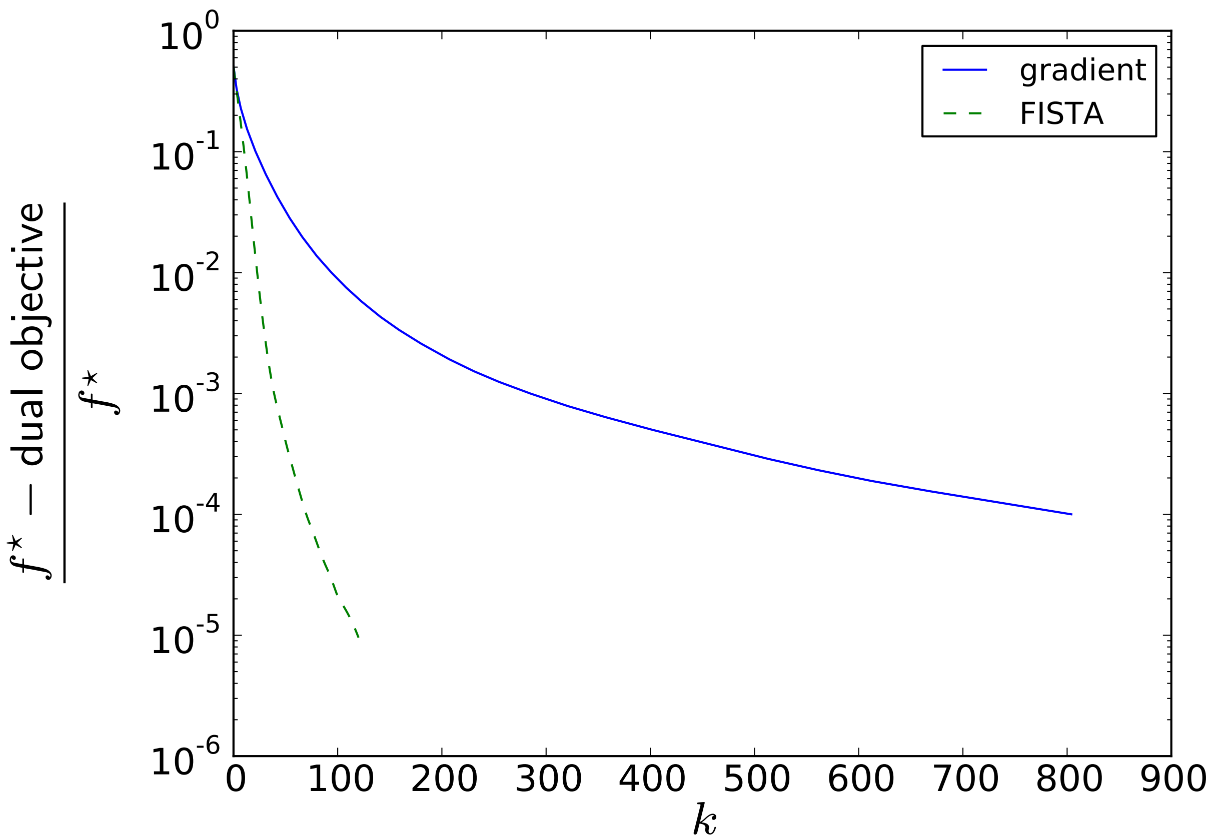
$$\begin{aligned} \hat{x} &= \operatorname*{argmin}_x \left(f(x) + \left(\sum_{i=1}^{p} B_i^T z_i\right)^T x \right) \\ z_i^+ &= P_{C_i}\left(z_i + tB_i\hat{x}\right), \quad i = 1, \ldots, p \end{aligned}$$

C_i is unit Euclidean norm ball in $\mathbf{R}^{m_i}$, if $B_i \in \mathbf{R}^{m_i \times n}$

numerical example

$$f(x) = \frac{1}{2}\|Cx - d\|_2^2$$

with randomly generated $C \in \mathbf{R}^{2000\times 1000}$, $B_i \in \mathbf{R}^{10\times 1000}$, $p = 500$



Minimization over intersection of convex sets

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C_1 \cap \cdots \cap C_m \end{array}$$

- f strongly convex; *e.g.*, $f(x) = \|x - a\|_2^2$ for projecting a on intersection
- sets C_i are closed, convex, and easy to project onto
- this is a special case of page 12-3 with g a sum of indicators

$$g(y_1, \ldots, y_m) = I_{C_1}(y_1) + \cdots + I_{C_m}(y_m), \qquad A = \left[\begin{array}{ccc} I & \cdots & I \end{array} \right]^T$$

dual proximal gradient update

$$\begin{aligned} \hat{x} &= \operatorname*{argmin}_x \left(f(x) + (z_i + \cdots + z_m)^T x \right) \\ z_i^+ &= z_i + t\hat{x} - tP_{C_i}(z_i/t + \hat{x}), \quad i = 1, \ldots, m \end{aligned}$$

Decomposition of separable problems

$$\text{minimize} \quad \sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} g_i(A_{i1}x_1 + \cdots + A_{in}x_n)$$

each f_i is strongly convex; g_i has inexpensive prox-operator

dual proximal gradient update

$$\begin{aligned} \hat{x}_j &= \operatorname*{argmin}_{x_j} \, (f_j(x_j) + \sum_{i=1}^{m} z_i^T A_{ij} x_j), \quad j = 1, \ldots, n \\ z_i^+ &= \operatorname{prox}_{tg_i^*}(z_i + t \sum_{j=1}^{n} A_{ij} \hat{x}_j), \quad i = 1, \ldots, m \end{aligned}$$

Outline

- proximal gradient method applied to the dual
- examples
- **alternating minimization method**

Primal problem with separable structure

composite problem with separable f

$$\text{minimize} \quad f_1(x_1) + f_2(x_2) + g(A_1x_1 + A_2x_2)$$

we assume f_1 strongly convex, but not necessarily f_2

dual problem

$$\text{maximize} \quad -f_1^*(-A_1^Tz) - f_2^*(-A_2^Tz) - g^*(z)$$

- first term is differentiable with Lipschitz continuous gradient
- prox-operator $h(z) = f_2^*(-A_2^Tz) + g^*(z)$ was discussed on page 10-10

Dual proximal gradient method

$$z^+ = \mathrm{prox}_{th}(z + tA_1 \nabla f_1^*(-A_1^T z))$$

- equivalent form using f_1:

$$z^+ = \mathrm{prox}_{th}(z + tA_1 \hat{x}_1) \quad \text{where } \hat{x}_1 = \underset{x_1}{\mathrm{argmin}} \, (f_1(x_1) + z^T A_1 x_1)$$

- from page 10-10, prox-operator of $h(z) = f_2^*(-A_2^T z) + g^*(z)$ is given by

$$\mathrm{prox}_{th}(w) = w + t(A_2 \hat{x}_2 - \hat{y})$$

where $\hat{x}_2$, $\hat{y}$ minimize an augmented Lagrangian

$$(\hat{x}_2, \hat{y}) = \underset{x_2, y}{\mathrm{argmin}} \, (f_2(x_2) + g(y) + \frac{t}{2} \|A_2 x_2 - y + w/t\|_2^2)$$

Alternating minimization method

starting at some initial z, repeat the following iteration

1. minimize the Lagrangian over x_1:

$$\hat{x}_1 = \underset{x_1}{\operatorname{argmin}} \left(f_1(x_1) + z^T A_1 x_1\right)$$

2. minimize the augmented Lagrangian over $\hat{x}_2$, $\hat{y}$:

$$(\hat{x}_2, \hat{y}) = \underset{x_2, y}{\operatorname{argmin}} \left(f_2(x_2) + g(y) + \frac{t}{2}\|A_1\hat{x}_1 + A_2 x_2 - y + z/t\|_2^2\right)$$

3. update dual variable:

$$z^+ = z + t(A_1\hat{x}_1 + A_2\hat{x}_2 - \hat{y})$$

Comparison with augmented Lagrangian method

augmented Lagrangian method (for problem on page 12-11)

1. compute minimizer $\hat{x}_1$, $\hat{x}_2$, $\hat{y}$ of the augmented Lagrangian

$$f_1(x_1) + f_2(x_2) + g(y) + \frac{t}{2} \|A_1 x_1 + A_2 x_2 - y + z/t\|_2^2$$

2. update dual variable:

$$z^+ = z + t(A_1 \hat{x}_1 + A_2 \hat{x}_2 - \hat{y})$$

differences with alternating minimization

- more general: AL method does not require strong convexity of f_1
- quadratic penalty in step 1 destroys separability

References

alternating minimization method

- P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control and Optimization (1991)
- P. Tseng, *Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming*, Mathematical Programming (1990)