12 Dual methods

In the last lectures, we described methods for the minimization of problems of the form $f(x)+h(x)$, where f is smooth, and h has a "simple" prox. In many situations however, one is faced with problems of the form:

$$\min_{x\in\mathbb{R}^n} \quad f(x) + h(Ax), \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$. Note that even if $\mathbf{prox}_h$ is easy to compute, computing $\mathbf{prox}_{h\circ A}$ can be hard.

Example (Signal denoising using total variation)**.** *Consider the problem of denoising a 1D signal $u \in \mathbb{R}^n$ with total-variation regularization*

$$\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^{n}(x_i - u_i)^2 + \lambda \sum_{i=1}^{n-1} |x_{i+1} - x_i|.$$

This problem can be put in the form (1) with $f(x) = \|x-u\|_2^2$, $h(x) = \|x\|_1$ and A is the discrete difference operator.

The main approach to deal with (1) is to use *duality*. We will then look at different algorithms one can apply to the dual problem:

- Dual subgradient method
- Dual proximal gradient method (if f is strongly convex)
- Dual proximal point method (augmented Lagrangian method)
- Alternating Direction Method of Multipliers (ADMM)

The dual problem We can rewrite problem (1) as

$$\min_{x,y} \quad f(x) + h(y) \quad \text{subject to} \quad y = Ax.$$

The Lagrangian is

$$L(x,y,z) = f(x) + h(y) + z^T(Ax - y) \tag{2}$$

and the dual function is

$$\begin{aligned} g(z) = \min_{x,y} L(x,y,z) &= \min_{x,y} f(x) + z^T Ax + h(y) - z^T y \\ &= \min_x \left\{f(x) + z^T Ax\right\} + \min_y \left\{h(y) - z^T y\right\} \\ &= -f^*(-A^T z) - h^*(z). \end{aligned} \tag{3}$$

where f^* and h^* are the *Fenchel conjugates* of f and h respectively, defined by

$$f^*(\xi) = \sup_x \left\{\langle \xi, x\rangle - f(x)\right\}$$

and similarly for h. So the dual problem is

$$\max_{z\in\mathbb{R}^n} -f^*(-A^T z) - h^*(z). \tag{4}$$

Fenchel conjugate Some properties about Fenchel conjugates will be useful. If f is lower semi-continuous (i.e., $\mathbf{epi}(f)$ is closed) then one can show that:

- Biduality: $f^{**} = f$
- $f(x) + f^*(\xi) = \langle \xi, x \rangle \iff x \in \partial f^*(\xi) \iff \xi \in \partial f(x)$
- If f is m-strongly convex, then $\mathbf{dom}(f^*) = \mathbb{R}^n$, f^* is smooth, and ∇f^* is $(1/m)$-Lipschitz.
- $\mathbf{prox}_{f^*}(x) = x - \mathbf{prox}_f(x)$.