1 Preliminaries and Notation

One such algorithm is the alternating direction method of multipliers [?], (ADMM). This algorithm solves problems of the form

$$\begin{aligned} \arg\min_{x} f(x) + g(z) \\ \text{s.t } Ux + Vz = c \end{aligned} \tag{1.0.1}$$

where f and g are assumed to be closed convex functions with range in $\mathbb{R}$, $U \in \mathbb{R}^{p\times n}$ and $V \in \mathbb{R}^{p\times m}$ are matrices (not assumed to have full rank), and $c \in \mathbb{R}^p$.

ADMM consists of iteratively minimising the augmented Lagrangian

$$L_p(x, z, \eta) = f(x) + g(z) + \eta^T (Ux + Vz - c) + \frac{\rho}{2}\|Ux + Vz - c\|_2^2$$

(η is a Lagrange multiplier), and ρ is a parameter we can choose to make $g(z)$ smooth [?], with the following iterations:

$$x^{k+1} := \arg\min_{x} L_\rho\left(x, z^k, \eta^k\right) \tag{1.0.2}$$

$$z^{k+1} := \arg\min_{z} L_\rho\left(x^{k+1}, z, \eta^k\right) \tag{1.0.3}$$

$$\eta^{k+1} := \eta^k + \rho\left(Ux^{k+1} + Vz^{k+1} - c\right) \tag{1.0.4}$$

The alternating minimisation works because of the decomposability of the objective function: the x minimisation step is independent of the z minimisation step and vice versa.

We will work frequently with the *dual* of problem (1.0.1) which is

$$D(\lambda) = \arg\max_{\lambda} -F^*\left(A^T x\right) + \langle\lambda, c\rangle - G^*\left(B^t\lambda\right) \tag{1.0.5}$$

where f^* is the convex conjugate:

$$F^*(p) = \sup_{p}\langle x, p\rangle - f(x) \tag{1.0.6}$$

For example the convex conjugate of the l_1 norm $g(x) = |x|$ is:

$$f^*(x^*) = \begin{cases} 0 & \text{if } |x| \le 1 \\ \infty & \text{if } |x| > 1 \end{cases} \tag{1.0.7}$$

One way to measure how far the iterations of ADMM are from the optimal solution is to define the primal and dual residuals:

$$r_k = c - Ax_k - Bz_k \tag{1.0.8}$$

$$d_k = \rho A^T B\left(x_k - x_{k-1}\right) \tag{1.0.9}$$

2 Convergence of ADMM

Theorem 2.1. *Consider the ADMM iterations defined by* (1.0.1). *Supposing that both f, g are strongly convex and that:*

$$\rho^3 \le \frac{\sigma_f \sigma_g}{\tau\left(A^T A\right)\tau\left(B^T B\right)} \tag{2.0.10}$$

Then the sequence of iterates $\{\eta_k\}$ *satisfy:*

$$D(\eta^*) - D(\eta) \leq \frac{||\eta^* - \eta_1||}{2\rho(k-1)} \tag{2.0.11}$$

Before we prove theorem 2.1 we need a couple of lemmas.

Definition 2.2.

$$\Psi(\lambda) = A\nabla F^*\left(A^T\lambda\right)$$
$$\Phi(\lambda) = B\nabla G^*\left(B^T\lambda\right)$$

Note that the derivative of $D(\lambda)$ is $b - \Phi - \Psi$, and maximising the dual problem is equivalent to finding a solution to $c \in \Phi(\lambda^*) + \Psi(\lambda^*)$. Since f, g are both strongly convex, then Ψ, Φ are both Lipchitz continuous with $L_\Psi \leq \frac{\tau(A^TA)}{\sigma_f}$ and $L_\Phi \leq \frac{\tau(B^TB)}{\sigma_g}$

Lemma 2.3. *Let* $\lambda, z \in \mathbb{R}^n$*, and define*

$$\lambda^{1/2} = \lambda + \rho\left(c - Ax^{k+1} - Bz^k\right) \tag{2.0.12}$$
$$\lambda^+ = \lambda + \left(c - Ax^{k+1} - Bz^{k+1}\right) \tag{2.0.13}$$

Then we have

$$Ax^{k+1} = A\nabla F^*\left(A^T\lambda^{1/2}\right) := \Psi\left(\lambda^{1/2}\right) \tag{2.0.14}$$
$$Bz^{k+1} = B\nabla G^*\left(B^T\lambda^+\right) := \Phi\left(\lambda^+\right) \tag{2.0.15}$$

Lemma 2.4. *Suppose that*

$$\rho^3 \leq \frac{\sigma_f \sigma_g}{\tau(A^TA)\,\tau(B^TB)} \tag{2.0.16}$$

and that $Bz = \Phi(\lambda)$*. Then for any* $\gamma \in \mathbb{R}^n$*:*

$$D\left(\lambda^+\right) - D(\gamma) \geq \rho^{-1}\langle\gamma - \lambda, \lambda - \lambda^+\rangle + \frac{1}{2\rho}\left|\left|\lambda - \lambda^+\right|\right|^2 \tag{2.0.17}$$

Proof

Theorem (2.1). ∎