

# 1 Preliminaries

Given a set of measurements of the form

$$y = Ax + n \quad (1.0.1)$$

where  $x \in \mathbb{R}^n$  is an  $s$ -sparse vector we wish to recover,  $y \in \mathbb{R}^m$  is a set of noisy measurements,  $A \in \mathbb{R}^{m \times n}$  is a design or measurement matrix s.t.  $x$  is not in the null-space of  $A$ , and  $z \in \mathbb{R}^m$  is AGWN. The signal  $x$  can be recovered by algorithms minimising the objective function:

$$L = \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1 \quad (1.0.2)$$

where  $\lambda$  is a parameter which trades off the reconstruction accuracy and sparsity of  $x$ : larger  $\lambda$  means sparser  $x$ .

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

$$\begin{aligned} \arg \min_x f(x) + g(z) \\ \text{s.t } Ux + Vz = c \end{aligned} \quad (1.0.3)$$

where  $f$  and  $g$  are assumed to be convex function 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

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

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

$$x^{k+1} := \arg \min_x L_p(x, z^k, \eta^k) \quad (1.0.4)$$

$$z^{k+1} := \arg \min_z L_p(x^{k+1}, z, \eta^k) \quad (1.0.5)$$

$$\eta^{k+1} := \eta^k + \rho (Ux^{k+1} + Vz^{k+1} - c) \quad (1.0.6)$$

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 illustrate an example, relevant to the type of problems encountered in signal processing.

ADMM can be formulated as an iterative MAP estimation procedure for the problem (1.0.2). We can write (1.0.2) in constrained form as:

$$\frac{1}{2} \|Ax - b\|_2^2 + \lambda \|z\|_1 \quad (1.0.7)$$

$$\text{s.t } z = x \quad (1.0.8)$$

i.e this is of the form (1.0.3) with  $f(x) = \|Ax - b\|_2^2$ ,  $g(z) = \lambda \|z\|_1$ ,  $U = I$ ,  $V = -I$ , and  $c = 0$ . The associated (augmented) Lagrangian is:

$$L_\rho = \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|z\|_1 + \eta(x - z) + \frac{\rho}{2} \|x - z\|^2 \quad (1.0.9)$$

The ADMM iterations for LASSO, which can be found by alternately differentiating (1.0.9) with respect to  $x, z$  and  $\eta$ , are (in closed form):

$$x^{k+1} := (A^T A + \rho I)^{-1} (A^T y + \rho (z^k - \eta^k / \rho)) \quad (1.0.10)$$

$$z^{k+1} := S_{\lambda/\rho} (x^{k+1} + \eta^k / \rho) \quad (1.0.11)$$

$$\eta^{k+1} := \eta^k + \rho (x^{k+1} - z^{k+1}) \quad (1.0.12)$$

where  $S_{\lambda/\rho}(\circ)$  is the soft thresholding operator:  $S_\gamma(x)_i = \text{sign}(x_i) (|x_i| - \gamma)^+$ . These can be found differentiating (1.0.9) with respect to  $x$  and  $z$  as follows:

$$\frac{\partial L}{\partial x} = -A^T (y - Ax) + \rho(x - z) + \eta$$

as

$$\frac{\partial}{\partial x} \|F(x)\|_2^2 = 2 \left( \frac{\partial}{\partial x} F(x) \right)^T F(x) \quad (1.0.13)$$

by the chain rule, and  $\partial/\partial x(Ax) = -A^T$  (see the Matrix Cookbook) as differentiation exchanges a linear operator with its adjoint.

Setting (1.0.13) to zero and collecting like terms:

$$(A^T A + \rho I) x = A^T y + \rho z - \eta \quad (1.0.14)$$

so we find the optimal  $x$  is:

$$x = (A^T A + \rho I)^{-1} (A^T y + \rho (z - \eta/\rho)) \quad (1.0.15)$$

note that this estimator is a weighted average of the ordinary least squares estimate  $(A^T y)$  and a Gaussian prior. This is to be expected, as the minimisation problem w.r.t  $x$  is an  $l_2$ -regularised MAP problem.

for  $z > 0$

$$\frac{\partial L}{\partial z} = \lambda + \rho(x - z) - \eta \quad (1.0.16)$$

from which we obtain:

$$z = x + \frac{1}{\rho}(\eta - \lambda)$$

since  $z > 0$  then  $x + \frac{1}{\rho}(\eta - \lambda I) > 0$  when  $x + \frac{\eta}{\rho} > \frac{\lambda}{\rho}$ . Similarly for  $z < 0$ :

$$\frac{\partial L}{\partial z} = -\lambda + \rho(x - z) \quad (1.0.17)$$

setting (1.0.17) to zero we obtain:

$$z = x + \frac{1}{\rho}(\eta + \lambda)$$

since  $z < 0$  then  $x + \frac{1}{\rho}(\eta + \lambda) < 0$  when  $x + \frac{\eta}{\rho} < -\frac{\lambda}{\rho}$ .  
at  $z = 0$  we find:

$$-\frac{\lambda}{\rho} \leq x + \frac{\eta}{\rho} \leq \frac{\lambda}{\rho}$$

i.e.

$$\left| x + \frac{\eta}{\rho} \right| \leq \frac{\lambda}{\rho} \quad (1.0.18)$$

combining (1.0.17), (1.0.16), (1.0.18) together we find the optimal  $z$  is:

$$z = \text{sign}\left(x + \frac{\eta}{\rho}\right) \max\left(\left| x + \frac{\eta}{\rho} \right| - \frac{\lambda}{\rho}, 0\right) \quad (1.0.19)$$

Together (1.0.15), (1.0.19) and the third step of (1.0.12) constitute the steps of the ADMM algorithm.