Closed forms for distributed LASSO?

Tom Kealy

February 12, 2015

This is a short note (follwing [1]), to write up some progress in seeing if the distributed ADMM version of LASSO (and also BPDN) have closed forms for their iterations.

1 LASSO-ADMM recap

The LASSO solves the following problem:

$$\arg\min x \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|x\|_1 \tag{1.0.1}$$

which is put into constrained form as:

$$\arg\min x \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|z\|_1 \text{ s.t } x-z=0 \tag{1.0.2}$$

The (augmented) Lagrangian of the problem is:

$$L_\rho(x,z,\eta) = \frac{1}{2}\|Ax-y\|_2^2 + \lambda\|z\|_1 + \eta(x-z) + \frac{\rho}{2}\|x-z\|_2^2 \tag{1.0.3}$$

which in turn leads to the following set of iterations:

$$x^{k+1} := \left(A^T A + \rho I\right)^{-1}\left(A^T b + \rho\left(z^k - \eta^k\right)\right) \tag{1.0.4}$$

$$z^{k+1} := S_{\lambda/\rho}\left(x^{k+1} + \eta^k/\rho\right) \tag{1.0.5}$$

$$\eta^{k+1} := \eta^k + \rho\left(x^{k+1} - z^{k+1}\right) \tag{1.0.6}$$

Where $S_{\lambda/\rho}(\circ)$ is the soft thresholding operator: $S_\gamma(x)_i = sign(x_i)\left(|x_i| - \gamma\right)^+$.

2 LASSO-ADMM on graphs

We have a network of nodes, which individually take some measurements according to the linear system:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{2.0.7}$$

where $\mathbf{A}$ is some measurement matrix satisfying the restricted isometry property (RIP) [2], $\mathbf{y}$ are the measurements taken by the nodes, $\mathbf{x}$ is the (sparse) signal we wish to reconstruct, and $\mathbf{n}$ is additive white Gaussian noise.

We model the network as an undirected graph $G=(V,E)$, where $V=\{1\ldots J\}$ is the set of vertices, and $E = V\times V$ is the set of edges. An edge between nodes i and j implies that the two nodes can communicate. The set of nodes node i can communicate with is written $\mathcal{N}_i$ and the degree of node i is $D_i = |\mathcal{N}_i|$.

Individually nodes make the following measurements:

$$\mathbf{y}_p = \mathbf{A}_p\mathbf{x} + \mathbf{n}_p \tag{2.0.8}$$

and the system (2.0.7) is formed by concatenating the individual nodes measurements together.

We assume that a proper (or approximate) colouring of the graph is available: that is each node is assigned a number from a set $C = \{1 \dots c\}$, and no node shares a colour with any neighbour.

To find the $\mathbf{x}$ we are seeking, to each node we give a copy of $\mathbf{x}, \mathbf{x}_p$ and we constrain the copies to be indentical across all edges in the network.

Specifically, we are solving:

$$\min_{\bar{x}=(x_1,\dots x_n)} \sum_J \|A_j x_j - y_j\|_2^2 + \frac{\lambda}{J}\|x\|_1 \tag{2.0.9}$$

$$\text{and } x_i = x_j \text{ if } \{i,j\} \in E \tag{2.0.10}$$

The paper [3] suggests that this prblem can be re-written as:

$$\min_{\bar{x}=(x_1,\dots x_n)} \sum_J \|A_j x_j - y_j\|_2^2 + \beta\|x_j\|_1 \tag{2.0.11}$$

$$\text{s.t. } \left(B^T \otimes I_n\right)\bar{x} = 0 \tag{2.0.12}$$

where $\beta = \frac{\lambda}{J}$, $\bar{x} = \left(x_1^T, \dots, x_J^T\right)^T$ which collects together J copies of a $n \times 1$ vector, and B is the arc-incidence matrix: a V by E matrix where each column is associated with an edge $(i,j) \in E$ and has 1 and -1 in the ith and jth entry respectively. Figures (2.1) and (2.2) show examples of a network and it's associated incidence matrix. Note that $\left(B^T \otimes I_n\right) \in \mathbb{R}R^{nE \times nJ}$.

We can also partition B and $\bar{x}$ by colour. $B = \left[B_1^T, \dots, B_C^T\right]^T$, $\bar{x} = \left(x_1^T, \dots, x_C^T\right)^T$. The constraints now require that $\sum_{c=1}^{C}\left(B_c^T \otimes I_n\right)\bar{x}_c = 0$

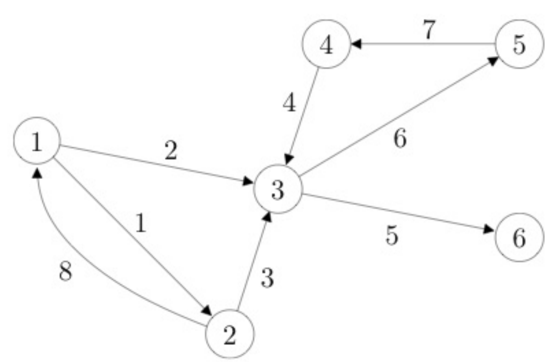

Figure 2.1: An example of a network

The Augmented Lagrangian for the problem (2.0.12) can be written down as:

$$L = \sum_{c=1}^{C} \|A_j x_j - y_j\|_2^2 + \beta\|x\|_1 + \eta^T\left(B_c^T \otimes I_n\right)\bar{x}_c + \frac{\rho}{2}\|\sum_{c=1}^{C}\left(B_c^T \otimes I_n\right)\bar{x}_c\|^2 \tag{2.0.13}$$

the general update (at nodes with colour 1) is:

$$
A = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & -1 & -1 & -1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0
\end{bmatrix}.
$$

Figure 2.2: The incidence matrix associated with 2.1

$$
\bar{x}_1^{k+1} = \sum_{p \in C_1} \|A_p x_p - y_p\|_2^2 + \beta \|x_p\|_1 + \eta^T \left(B_p^T \otimes I_n\right) \bar{x}_p + \frac{\rho}{2} \| \left(B_p^T \otimes I_n\right) \bar{x}_1 + \sum_{c=2}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c \|^2 \tag{2.0.14}
$$

The paper [3, P.3] shows how (2.0.14) can be written in the following form

$$
\bar{x}_1^{k+1} = \arg\min_x \sum_{j \in C_1} \|A_1 x_1 - y_1\|_2^2 + \beta \|x_1\|_1 + \left(\sum_{k \in \mathcal{N}_1} sign\left(k-1\right) \eta_{\{1,k\}} - \rho x_k \right)^T x_1 + \frac{\rho}{2} D_i \left\|x_1\right\|^2 \tag{2.0.15}
$$

where D_p is the degree of node p, and C_1 is the set of nodes all of the same colour.

We should quickly note that this is an instance of the extended-ADMM algorithm:

$$
\begin{aligned}
& \underset{x}{\text{minimize}} && f_1(x_1) + f_2(x_2) + \ldots + f_q(x_q) \\
& \text{subject to} && U_1 x_1 + U_2 x_2 + \ldots U_q x_q = 0.
\end{aligned}
$$

with $f_r = \|A_r x_r - y_r\|_2^2 + \beta \|x_r\|^1$, $U_r = \left(B_r^T \otimes I_n\right)$, and x_r is a local copy of x $\forall r = 1, \ldots, C$.

The goal of this note is to introduce a set of dummy variables z_r, to try to solve a problem which looks like:

$$
\begin{aligned}
& \underset{x}{\text{minimize}} && g_1(x_1) + h_1(z_1) + \ldots + g_q(x_q) + h_q(x_2) \\
& \text{subject to} && U_1 x_1 + V_1 z_1 \ldots U_q x_q + V_q z_q = 0.
\end{aligned}
$$

with $g_r = \|A_r x_r - y_r\|_2^2$, $h_r = \beta \|x_r\|^1$, $U_r = \left(B_r^T \otimes I_n\right)$, $V_r = -I$ and x_r, z_r are a local copies of x, z $\forall r = 1, \ldots, C$. We seek a set of closed form iterations for the minimisation (2.0.15), as (2.0.15) cannot be done without an auxiliary minimisation algorithm.

The re-writing hinges on re-writing the the second and third terms of (2.0.14).

The third term can be written as:

$$
\frac{\rho}{2} \| \left(B_1^T \otimes I_n\right) \bar{x}_1 + \sum_{c=2}^{C} \left(B_1^T \otimes I_n\right) \bar{x}_j \|^2 =
$$

$$
\frac{\rho}{2} \bar{x}_1^T \left(B_1^T \otimes I_n\right)^T \left(B_1^T \otimes I_n\right) \bar{x}_1 + \rho \bar{x}_1^T \sum_{c=2}^{C} \left(B_1^T \otimes I_n\right)^T \left(B_c^T \otimes I_n\right) \bar{x}_c + \| \sum_{c=2}^{C} \left(B_c^T \otimes I_n\right) \bar{x}_c \|^2
$$

We will investigate the first and second terms more, as the last term doesn't depend on $\bar{x}_1$ and can be dropped.

Lemma 2.1.
$$\frac{\rho}{2}\bar{x}_1^T\left(B_1^T\otimes I_n\right)^T\left(B_1^T\otimes I_n\right)\bar{x}_1=\frac{\rho}{2}\sum_{l\in C_1}D_l\|x_l\|^2 \tag{2.0.16}$$

Proof. $\left(B_1^T\otimes I_n\right)^T\left(B_1^T\otimes I_n\right)=B_1B_1^T\otimes I_n$ from taking the transpose of the first term, and using the properties of the Kronecker product. BB^T is a $J\times J$ matrix, with the degree of the nodes on the main diagonal and -1 in position (i,j) if nodes i and j are neighbours (i.e BB^T is the graph Laplacian). The trace of $B_1B_1^T$ is simply the sum of the degrees of nodes with colour 1. □

Lemma 2.2. $\rho\bar{x}_1^T\sum_{c=2}^{C}\left(B_1^T\otimes I_n\right)^T\left(B_c^T\otimes I_n\right)\bar{x}_c=-\rho\sum_{l\in C_1}\sum_{m\in N_l}x_l^Tx_m^k$

Proof. $\left(B_1^T\otimes I_n\right)^T\left(B_c^T\otimes I_n\right)=B_1B_c^T\otimes I_n$ by repeating the steps from lemma 2.1. $B_1B_c^T$ corresponds to an off diagonal block of the graph Laplacian, and so counts how many neighbours each node with colour 1 has. □

Finally, we need to consider the second term from (2.0.14):

$$\eta^T\left(B_1^T\otimes I_n\right)\bar{x}_1=\sum_{l\in C_1}\sum_{m\in N_l}sign\left(m-l\right)\eta_{ml}^Tx_l \tag{2.0.17}$$

where η is decomposed edge-wise: $\eta=(\ldots,\eta_{ij},\ldots)$, such that $\eta_{i,j}=\eta_{j,i}$ and is associated with the constraint $x_i=x_j$.

adding together this with the two lemmas, lets us write (2.0.14) as (2.0.15).

To tidy (2.0.15) up define:

$$\nu_i=\left(\sum_{k\in\mathcal{N}_i}sign\left(k-i\right)\eta_{\{i,k\}}-\rho x_k\right) \tag{2.0.18}$$

this is a rescaled version of the Lagrange multiplier, η, which respects the graph structure. Finally (2.0.14) reduces to:

$$\bar{x}_1^{k+1}=\sum_{j\in C_1}\|A_1x_1-y_1\|_2^2+\|x_1\|_1+\nu_1^Tx_1+\frac{\rho}{2}D_i\left|\left|x_1\right|\right|^2 \tag{2.0.19}$$

We seek a set of closed form iterates, like (1.0.6) except now including information about the network.

To this end we can write (2.0.12) as:

$$\min_{\bar{x}=(x_1,\ldots x_n)}\sum_J\|A_jx_j-y_j\|_2^2+\|z_j\|_1 \tag{2.0.20}$$
$$\text{s.t. }\left(B^T\otimes I_n\right)\bar{x}=0 \tag{2.0.21}$$
$$\text{and }\bar{x}-\bar{z}=0 \tag{2.0.22}$$

which can be written more compactly as:

$$\min_{\bar{x}=(x_1,\ldots x_n)}\sum_J\|A_jx_j-y_j\|_2^2+\|z_j\|_1 \tag{2.0.23}$$
$$\text{s.t. }\left(\left(B^T\otimes I_n\right)+I_n\right)\bar{x}-\bar{z}=0 \tag{2.0.24}$$

The Augmented Lagrangian for this problem can be written down as:

$$L = \sum_J \|A_j x_j - y_j\|_2^2 + \|x_j\|_1 + \eta^T M\bar{x} - \bar{z} + \frac{\rho}{2}\|M\bar{x} - \bar{z}\|^2 \tag{2.0.25}$$

where we have defined $M = \left(\left(B^T \otimes I_n\right) + I_n\right)$

Conjecture:

Following steps similar to [3] this Lagrangian can be written in the following form

$$L = \sum_J \|A_j x_j - y_j\|_2^2 + \|z_j\|_1 + \nu_j^T \left(x_j - z_j\right) + \frac{\rho}{2} D_i \left|\left|x_j - z_j\right|\right|^2 \tag{2.0.26}$$

Then by differentiating with respect to x_j and z_j we can find closed forms for the updates as:

$$x_j^{k+1} := \left(A_j^T A_j + \rho D_J I\right)^{-1} \left(A_j^T y_j + \rho D_j z^k - \nu^k\right) \tag{2.0.27}$$

$$z^{k+1} := S_{\lambda/J}\left(x_j^{k+1} + \frac{1}{\rho D_j}\nu_j^{k+1}\right) \tag{2.0.28}$$

$$\nu^{k+1} := \nu^k + \rho\left(x^{k+1} - z^{k+1}\right) \tag{2.0.29}$$

As a step towards this, we need to work out what $\bar{x}_1^T M_1^T M_1 \bar{x}_1$ and $\bar{x}_1^T \sum_{c=2}^{C} M_1^T M_c \bar{x}_c$ is in our case.

$$\begin{aligned}
&\bar{x}_1^T \left(\left(B^T \otimes I_n\right) + I_n\right)^T \left(\left(B^T \otimes I_n\right) + I_n\right) \bar{x}_1 = \\
&\bar{x}_1 \left(B_1 B_1^T \otimes I_n + I_n^T \left(B_1^T \otimes I_n\right) + \left(B_1^T \otimes I_n\right)^T I_n + I_n\right) \bar{x}_1
\end{aligned}$$

we know what $\bar{x}_1^T \left(B_1 B_1^T \otimes I_n\right) \bar{x}_1$ is, from lemma 2.1. The second and third terms $\bar{x}_1^T \left(B_1^T \otimes I_n\right) \bar{x}_1$ and $\bar{x}_1^T \left(B_1^T \otimes I_n\right)^T \bar{x}_1$ will equal zero because $\left(B_1^T \otimes I_n\right) \bar{x}_1 = 0$ from the constraints of the problem.

Also,

$$\begin{aligned}
&\rho \bar{x}_1^T \sum_{c=2}^{C} \left(\left(B_1^T \otimes I_n\right) + I_n\right)^T \left(\left(B_c^T \otimes I_n\right) + I_n\right) \bar{x}_c = \\
&\rho \bar{x}_1^T \sum_{c=1}^{C} \left(\left(B_1 B_c^T \otimes I_n\right) + \left(B_c^T \otimes I_n\right) I_n^T + I_n \left(B_1^T \otimes I_n\right)^T + I_n\right) \bar{x}_c \\
&= -\rho \sum_{l \in C_1} \sum_{m \in N_l} x_l^T x_m^k + \rho \sum_{c=2}^{C} \bar{x}_1^T \bar{x}_c
\end{aligned}$$

using lemma 2.2, and again the middle terms are equal to 0.

The final term to deal with is:

$$\eta^T \left(\left(B_1^T \otimes I_n\right) + I_*\right) = \eta^T \left(B_1^T \otimes I_n\right) \bar{x}_1 + \eta^T I_8 \bar{x}_1 \tag{2.0.30}$$

the first term we have dealt with previously in (2.0.17).

Collecting all terms together we now have:

$$\eta^T \left(\left(B_1^T \otimes I_n\right) + I_*\right) + \| \sum_{c=1}^{C} M_c \bar{x}_c \|^2 = \tag{2.0.31}$$

$$\eta^T \left(B_1^T \otimes I_n\right) \bar{x}_1 + \eta^T I_8 \bar{x}_1 + -\rho \sum_{l \in C_1} \sum_{m \in N_l} x_l^T x_m^k + \rho \sum_{c=2}^{C} \bar{x}_1^T \bar{x}_c + B_1 B_1^T \otimes I_n + \|x_1\|^2 \tag{2.0.32}$$

so, collecting all terms together we have:

$$L = \sum l \in c_1 \|A_l x_l - y_l\|_2^2 + \beta \|z_l\|_1 + \nu_l^T \bar{x}_1 + \frac{\rho D_l}{2} \|x_l\|^2 + \|x_l\|^2 + \rho \sum_{c=2}^{C} x_l^T x_c + \eta^T I_* \tag{2.0.33}$$

collecting like terms together we have:

$$L = \sum l \in c_1 \|A_l x_l - y_l\|_2^2 + \beta \|z_l\|_1 + \mu^T \bar{x}_l + \gamma \|x_l\|^2 \tag{2.0.34}$$

where

$$\mu = \sum_{m \in N_l} \left(sign\left(m - l\right) + 1\right) \eta^T - 2\rho \bar{x}_m \tag{2.0.35}$$

and

$$\gamma = 1 + \frac{\rho D_l}{2} \tag{2.0.36}$$

3 Second Attempt

Define

$$v := \left(\left(B^T \otimes I_n\right) + I\right) \bar{x} - \bar{z} = \sum_{j=1}^{J} \left[\left(B^T e_j\right) \otimes x_j - e_j \otimes z_j\right] \tag{3.0.37}$$

where we have defined $\bar{x} = \sum_{j=1}^{J} e_j \otimes x_j$ and e_j is a $J \times 1$ unit vector.
we are trying to calculate $v^T v$.

$$\begin{aligned} v^T v &= \sum_{k=1}^{J} \sum_{j=1}^{J} \left[\left(e_k^T B\right) \otimes x_k^T - e_k^T \otimes z_k\right] \left[\left(B^T w_c\right) \otimes \bar{x}_j - e_j \otimes z_j\right] \\ &= \sum_{k,j} \left(e_k^T B B^T e_j \otimes x_k^T x_j\right) - \left(e_k^T B e_j \otimes x_k^T z_j\right) - \left(e_k^T B^T e_j \otimes z_k^T x_j\right) + \left(e_k^T e_j \otimes z_k^T z_j\right) \end{aligned}$$

Now $\left(e_k^T B^T e_j \otimes z_k^T x_j\right) = 0$ as the first term $e_k^T B^T e_j$ is an inner product between a $J \times 1$vector containing a single 1 and a single -1 and a $1 \times J$ vector containing all $1s$. Similar reasoning applies to the other cross term.

So:

$$v^T v = \sum_{k,j} \left(e_k^T B B^T e_j \otimes x_k^T x_j\right) + \left(e_k^T e_j \otimes z_k^T z_j\right)$$

we can try to solve the same problem with a smaller parameter space: write $c(i)$ for the colour of the i^{th} node, and w_c for the vector identifying the nodes of colour c. Then we can equivalently write

$$\bar{x} = \sum_{c=1}^{C} w_c \otimes x_c \tag{3.0.38}$$

as before, define

$$v := v := \left(\left(B^T \otimes I_n\right) + I\right) \bar{x} - \bar{z} = \sum_{c=1}^{C} \left[\left(B^T w_c\right) \otimes x_c - w_c \otimes z_c\right] \tag{3.0.39}$$

$$\begin{aligned} v^T v &= \sum_{e=1}^{C} \sum_{c=1}^{C} \left[\left(w_e^T B\right) \otimes x_e^T - w_e^T \otimes z_e^T\right] \left[\left(B^T w_c\right) \otimes x_c - w_c \otimes z_c\right] \\ &= \sum_{e,c} \left(w_e^T B B^T w_c \otimes x_e^T x_c\right) - \left(w_e^T B w_c \otimes x_e^T z_c\right) - \left(w_e^T B^T w_c \otimes z_e^T x_c\right) + \left(w_e^T w_c \otimes z_e^T z_c\right) \end{aligned}$$

4 Third Attempt

Earlier today (February 12, 2015), we wrote down two Lagrangians as there was some confusion (on my end) about which one would be preferable. They were:

$$L = \sum_J \|A_j x_j - y_j\|_2^2 + \|z_j\|_1 + \eta^T M \bar{x} + \theta^T \left(\bar{x} - \bar{z}\right) + \frac{\rho}{2} \|M\bar{x} + \bar{x} - \bar{z}\|^2 \tag{4.0.40}$$

where $M = \left(B^T \otimes I_n\right)$, and

$$L = \sum_J \|A_j x_j - y_j\|_2^2 + \|z_j\|_1 + \eta^T M \bar{x} + \theta^T \left(\bar{x} - \bar{z}\right) + \frac{\rho}{2} \|M\bar{x}\|^2 + \frac{\rho}{2} \|\bar{x} - \bar{z}\|^2 \tag{4.0.41}$$

The second (4.0.41) Lagrangian is correct, as $Mx + (x - z)$ isn't legal. This is fortuitous, as we can use the heavy lifting from [3] to write the Lagrangian as:

$$L = \sum_J \|A_j x_j - y_j\|_2^2 + \|z_j\|_1 + \nu^T \bar{x} + \theta^T \left(\bar{x} - \bar{z}\right) + \frac{\rho D_j}{2} \|\bar{x}_j\|^2 + \frac{\rho}{2} \|\bar{x}_j - \bar{z}_j\|^2 \tag{4.0.42}$$

References

[1] Stephen Boyd. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2010.

[2] Emmanuel J Candes, Justin Romberg, and Terence Tao. Robust Uncertainty Principles : Exact Signal Frequency Information. 52(2):489–509, 2006.

[3] João FC Mota, João MF Xavier, Pedro MQ Aguiar, and Markus Puschel. D-admm: A communication-efficient distributed algorithm for separable optimization. *Signal Processing, IEEE Transactions on*, 61(10):2718–2723.