**Fused LASSO using alternative direction method of multipliers (ADMM)**

In many biological applications, we are interested in the association of a response variable $Y \in \mathbb{R}^{N\times 1}$, with respect to some predictor variables $X \in \mathbb{R}^{N\times p}$, across varying conditions. In the simple 2-condition scenario, we have the following model:

$$Y = X\beta + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

$$Y = [Y_1, Y_0]^T$$

$$X = BlockDiag(X_1, X_0)$$

In order to choose relevant predictor variables, we can impose an L1 penalty on the beta coefficient vector, $\beta$. Additionally, when we expect the relationship between the predictors and response to be mostly the same across groups, it is useful to impose a penalty on the beta coefficients of paired coefficients, $\beta_j$ *and* $\beta_{j+p}$, between treatment and control groups.

$$argmin_\beta \ \frac{1}{2N}\|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|D\beta\|_1$$

However, because there are multiple nondifferentiable L1 penalty factors, the objective function cannot be optimized using standard gradient based methods. To efficiently solve this, we introduce two auxiliary variables, z and d, in place of $\beta$ and $D\beta$.

$$z = \beta$$

$$d = D\beta$$

Our new objective function becomes:

$$argmin_\beta \ \frac{1}{2N}\|Y - X\beta\|_2^2 + \lambda_1\|z\|_1 + \lambda_2\|d\|_1$$

$$s.t. \ \ \beta - z = 0; \quad D\beta - d = 0$$

As this is a constrained optimization problem, we use the augmented Lagrangian as our objective function (See appendix for details on the derivation):

$$\mathcal{L}(\beta, z, d, u, w) = \frac{1}{2N}\|Y - X\beta\|_2^2 + \lambda_1\|z\|_1 + \lambda_2\|d\|_1 + \frac{\rho_1}{2}\|\beta - z + \mu\| + \frac{\rho_2}{2}\|D\beta - d + w\|$$

We can now differentiate with respect to $\beta$ and then update $z$ and $d$ at each step. This procedure resembles a block coordinate descent algorithm, where we alternate between optimizing over $\beta$, $(z, d)$, and updating the dual variables $(u, w)$. The key is that we have removed the non-differentiable terms with respect to $\beta$ and exchanged them for auxiliary variables so that the function is differentiable over $\beta$.

---

**Step 1 – Update $\boldsymbol{\beta}$**

To update $\beta$, we differentiate the augmented Lagrangian with respect to $\beta$ and set it equal to zero. This results in the following linear system:

$$\left(\frac{1}{N}X^TX + \rho_1 I + \rho_2 D^T D\right)\beta^{(k+1)} = \frac{1}{N}X^TY + \rho_1\left(z^{(k)} - \mu^{(k)}\right) + \rho_2\left(z^{(k)} - \mu^{(k)}\right) + \rho_2 D^T(d^{(k)} - w^{(k)})$$

Which we solve by Cholesky factorization of the LHS and backsubstitution.

---

**Step 2 – Update $\boldsymbol{z}$ and $\boldsymbol{d}$**

At the current step $k$, $z^{(k+1)}$ and $d^{(k+1)}$ can be updated as:

$$z^{(k+1)} = S\left(\beta^{(k+1)} + \mu^{(k)}, \frac{\lambda_1}{\rho_1}\right)$$
$$d^{(k+1)} = S\left(D\beta^{(k+1)} + w^{(k)}, \frac{\lambda_2}{\rho_2}\right)$$

Where $S(x, t)$ is the soft threshold operator:

$$S(x, t) = \operatorname{sign}(x) \cdot \max(|x| - t, 0)$$

The threshold $t_j = \frac{\lambda_j}{\rho_j}$ is based on the Karush Kuhn Tucker conditions of the augmented Lagrangian (see appendix for details).

**Step 3 – Update u, w,**

$$u^{(k+1)} = \mu^{(k)} + \beta^{(k+1)} - z^{(k+1)}$$

$$w^{(k+1)} = w^{(k)} + D\beta^{(k+1)} - d^{(k+1)}$$

We iterate through steps 1-3 until the augmented Lagrangian decreases by less than $10^{-5}$ for ten consecutive iterations.

**Appendix**