

EE364b Spring 2020 Homework 7

Due Friday 5/29 at 11:59pm via Gradescope

7.1 (7 points) ADMM and Proximal Methods for Group Lasso

Consider a regression problem with a data matrix $X \in \mathbb{R}^{n \times (p+1)}$, where each column represents a predictor. Suppose that the matrix X is split into J groups over its columns:

$$X = [\vec{1} \ X_{(1)} \ X_{(2)} \ \dots \ X_{(J)}]$$

where $\vec{1} = [1 \ 1 \ \dots \ 1] \in \mathbb{R}^n$ is a vector of all ones. The groups are typically determined by the types of predictors. To achieve sparsity over the groups rather than individual predictors, we may write $\beta = (\beta_0, \beta_{(1)}, \dots, \beta_{(J)})$, where β_0 is an intercept term and each $\beta_{(j)}$ is an appropriate coefficient block of β corresponding to $X_{(j)}$, and solve the regularized optimization problem:

$$\min_{\beta \in \mathbb{R}^{p+1}} f(\beta) + h(\beta).$$

Here $h(\beta)$ is a convex regularization term to promote the sparsity over groups. In this problem, we will use group Lasso to predict the Parkinson's disease (PD) symptom score on the Parkinsons dataset¹. The PD symptom score is measured on the unified Parkinson's disease rating scale (UPDRS). This data contains 5,785 observations, 18 predictors (provided in **X_train.csv**), and an outcome - the total UPDRS (provided in **y_train.csv**)

The 18 columns in the predictor matrix have the following groupings (in column ordering):

- age: Subject age in years
- sex: Subject gender, 0-male, 1-female
- Jitter(%), Jitter(Abs), Jitter:RAP, Jitter:PPQ5, Jitter:DDP: Several measures of variation in fundamental frequency of voice
- Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA: Several measures of variation in amplitude of voice
- NHR, HNR: Two measures of ratio of noise to tonal components in the voice
- RPDE: A nonlinear dynamical complexity measure
- DFA: Signal fractal scaling exponent
- PPE: A nonlinear measure of fundamental frequency variation

¹'Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. BioMedical Engineering Online 2007, 6:23 (26 June 2007)

We consider the group LASSO problem, where $h(\beta) = \lambda \sum_j w_j \|\beta_{(j)}\|_2$:

$$\min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{2n} \|X\beta - y\|_2^2 + \lambda \sum_j w_j \|\beta_{(j)}\|_2$$

A typical choice for weights on groups w_j is $\sqrt{p_j}$, where p_j is number of predictors that belong to the j th group, to account for the group sizes. We will solve the problem using both ADMM and proximal gradient descent method.

- (a) (1 point) Derive the proximal operator for the convex function $h(\beta) = \lambda \sum_j w_j \|\beta_{(j)}\|_2$.
- (b) (1 point) Derive the ADMM update for the objective as described in the lecture slides (page 17, link).
Hint: You can let $h(\alpha) = \lambda \sum_j w_j \|\alpha_{(j)}\|_2$, and rewrite the original objective as $f(\beta) + h(\alpha)$ with the consensus constraint $\alpha = \beta$.
- (c) (1 point) Implement ADMM to solve the least squares group lasso problem on the Parkinsons dataset. Set $\lambda = 0.02$.
- (d) (1 point) Derive the proximal gradient method updates for the objective as described in the lecture slides (page 20, link).
- (e) (1 point) Implement proximal gradient descent to solve the least squares group lasso problem on the Parkinsons dataset. Set $\lambda = 0.02$. Use a fixed step-size $t = 0.005$.
- (f) (1 point) Plot $f^k - f^*$ versus k for the first 10000 iterations on a semi-log scale for both methods for the training data, where f^k denotes the objective value at step k , and the optimal objective value is $f^* = 49.9649$. Print the components of the solutions numerically. Which groups are selected, i.e., non-zero at the solution?
- (g) (1 point) (Extra credit) Implement accelerated proximal gradient method described in lecture slides (page 22, link) and compare with the results in (f).

7.2 (6 points) *ADMM for smart grid device coordination.* We consider an electrical grid consisting of N devices that exchange electricity over T time periods. Device i has energy profile $p^i \in \mathbf{R}^T$, with p_t^i denoting the energy consumed by device i in time period t , for $t = 1, \dots, T$, $i = 1, \dots, N$. (When $p_t^i < 0$, device i is producing energy in time period t .) Each device has a convex objective function $f_i : \mathbf{R}^T \rightarrow \mathbf{R}$, which we also use to encode constraints, by setting $f_i(p^i) = \infty$ for profiles that violate the constraints of device i . In each time period the energy flow has to balance, which means

$$\sum_{i=1}^N p_t^i = 0, \quad t = 1, \dots, T.$$

The optimal profile coordination problem is to minimize the total cost, $\sum_{i=1}^N f_i(p^i)$, subject to the balance constraint, with variables p^i , $i = 1, \dots, N$.

In this problem you will use ADMM to solve the optimal profile coordination problem in a distributed way, with each device optimizing its own profile, and exchanging messages to coordinate all of the profiles.

From this point on, we consider a specific (and small) problem instance. There are three devices: a generator, a fixed load, and a battery, with cost functions described below.

- *Generator.* The generator has upper and lower generator limits: $P^{\min} \leq -p_t \leq P^{\max}$, for $t = 1, \dots, T$. (Note the minus sign, since a generator's profile is typically negative, using our convention.) The objective is

$$f_{\text{gen}}(p) = \sum_{t=1}^T (\alpha(-p_t)^2 + \beta(-p_t)),$$

where $\alpha, \beta > 0$ are given constants.

- *Fixed load.* The fixed load has zero objective function and the constraint that its power profile must equal a given consumption profile $d \in \mathbf{R}^T$.
 - *Battery.* The battery has zero objective function, and charge/discharge limits given by C and D , respectively: $-D \leq p_t \leq C$, for $t = 1, \dots, T$. The battery is initially uncharged (*i.e.*, $q_1 = 0$), so its charge level in period t is $q_t = \sum_{\tau=1}^{t-1} p_\tau$ (we neglect losses for this problem). The charge level must be nonnegative, and cannot exceed the battery capacity: $0 \leq q_t \leq Q$, $t = 1, \dots, T+1$. The charge level is extended to time $T+1$ to allow the battery to charge/discharge in time T , subject to the operational constraints.
- (a) (3 points) Use CVX/CVXPY to solve the problem with data given in `admm_smart_grid_data.m`. Plot the (optimal) power profile for the generator and battery, as well as the battery charge level.
- (b) (3 points) Implement ADMM for this problem (you may use CVX/CVXPY to solve each device's local optimization in each ADMM iteration). Experiment with a few values of the parameter ρ to see its effect on the convergence rate of the algorithm. Plot the norm of the energy balance residual, versus iteration. Plot the power profiles of the generator and battery, as well as the energy balance residual, for several values of iteration (say, after one iteration, after 10 iterations, and after 50). Check the results against the solution found by CVX/CVXPY.