Math 6397
Computational and Mathematical Methods in Data Science
Problem Set 1
Due on Friday, March 1, at 10 PM

1 Assignments

1. We consider the ℓ^1-regularized linear regression problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{x}\|_1 \tag{1}$$

where $\alpha > 0$ is a scalar regularization parameter. We assume that the $m \times n$ matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{m,n}$ is of size $m = 1\,500$ (examples) and $n = 5\,000$ (features). We choose $a_{ij} \sim \mathcal{N}(0,1)$. We normalize the columns of $\mathbf{A}$ to have unit ℓ_2-norm. A "true" value $\mathbf{x}_{\text{true}} \in \mathbb{R}^n$ is generated with 100 nonzero entries, each sampled from a $\mathcal{N}(0,1)$ distribution. The associated labels $\mathbf{y} \in \mathbb{R}^m$ are computed as $\mathbf{y} = \mathbf{A}\mathbf{x}_{\text{true}} + \boldsymbol{\eta}$, where $\eta \sim \mathcal{N}(\mathbf{0}, 1\text{e–}3\mathbf{I})$. This corresponds to a signal-to-noise ratio $\mathbf{A}\mathbf{x}_{\text{true}}\|_2^2/\|\boldsymbol{\eta}\|_2^2$ of about 60.

The ADMM form of this problem can be written as

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{subject to} \quad & \mathbf{x} - \mathbf{z} = 0 \end{aligned}$$

where $f(\mathbf{x}) = (1/2)\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$ and $g(\mathbf{z}) = \alpha\|\mathbf{z}\|_1$. The augmented Lagrangian of the problem above is given by

$$\begin{aligned} \ell_\rho(\mathbf{x}, \boldsymbol{\nu}) &= f(\mathbf{x}) + g(\mathbf{z}) + \boldsymbol{\nu}^{\mathsf{T}}(\mathbf{x} - \mathbf{z}) + \rho\|\mathbf{x} - \mathbf{z}\|_2^2 \\ &= \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{z}\|_1 + \frac{\rho}{2}\|\mathbf{x} - \mathbf{z} + \mathbf{u}\|_2^2 \end{aligned}$$

with $\mathbf{u} = \boldsymbol{\nu}/\rho$.

We summarize the ADMM steps in Alg. 1. We note that $\mathbf{A}^{\mathsf{T}}\mathbf{A} + \rho\mathbf{I} \succ \mathbf{0}$ for $\rho > 0$. Notice that the $\mathbf{x}$ update is essentially a ridge regression. You can solve the associated linear system using a direct solver. The operator S_κ in Alg. 1 is given by $S_\kappa(w) = (w - \kappa)_+ - (-a - \kappa)_+$.

Algorithm 1 ADMM steps for solving (1).

1: $\mathbf{x}^{(k+1)} \leftarrow (\mathbf{A}^{\mathsf{T}}\mathbf{A} + \rho\mathbf{I})^{-1}(\mathbf{A}^{\mathsf{T}}\mathbf{y} + \rho(\mathbf{z}^{(k)} - \mathbf{u}^{(k)}))$
2: $\mathbf{z}^{(k+1)} \leftarrow S_{\alpha/\rho}(\mathbf{x}^{(k+1)} + \mathbf{u}^{(k)})$
3: $\mathbf{u}^{(k+1)} \leftarrow \mathbf{u}^{(k)} + \mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)}$

We terminate the solver if the primal residual $\mathbf{r}^{(k)}$ and the dual residual $\mathbf{s}^{(k)}$ are below given tolerances ϵ_{pri} and ϵ_{dual}, i.e.,

$$\|\mathbf{r}^{(k)}\|_2^2 \le \epsilon_{\text{pri}} \quad \text{and} \quad \|\mathbf{s}^{(k)}\|_2^2 \le \epsilon_{\text{dual}}.$$

The primal residual is given by $\mathbf{r}^{(k+1)} = \mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)}$ and the dual residual is given by $\mathbf{s}^{(k+1)} = -\rho(\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)})$, respectively. The tolerances for the primal and dual residual can be computed based on

$$\begin{aligned}\epsilon_{\text{pri}} &= \sqrt{n}\epsilon_{\text{abs}} + \epsilon_{\text{rel}} \max\{\|\mathbf{x}^{(k)}\|_2, \| - \mathbf{z}^{(k)}\|_2\} \\ \epsilon_{\text{dual}} &= \sqrt{n}\epsilon_{\text{abs}} + \epsilon_{\text{rel}}\|\rho\mathbf{u}^{(k)}\|_2\end{aligned}$$

We select $\rho = 1$ and use the tolerances $\epsilon_{\text{abs}} = 1\text{e–}4$ and $\epsilon_{\text{rel}} = 1\text{e–}2$. We initialize the variables $\mathbf{u}^{(0)}$ and $\mathbf{z}^{(0)}$ as $\mathbf{0}$. The regularization parameter α is set to $\alpha = \|\mathbf{A}^\mathsf{T}\mathbf{y}\|_\infty/10$.

a) Derive the steps of the ADMM algorithm in Alg. 1

b) Implement the ADMM algorithm