Algorithms for ℓ_1 minimization

Yihang Chen
Department of Mathematics
Peking University
1700010780

1 Problem formulation

Lasso

Consider the ℓ_1-regularized problem

$$\min_x \quad f_\mu(x) = g(x) + \mu h(x) := \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1 \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $\mu > 0$ are given.

2 CVX solutions

We can obtain and solutions directly from cvx mosek and cvx gurobi.

The ℓ_1 regularized problem (1.1) is equivalent to the following optimization problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|A(x^+ - x^-) - b\|_2^2 + \mu\mathbb{1}^\top(x^+ + x^-) \\ s.t. \quad & x^+, x^- \geq 0 \end{aligned} \tag{2.1}$$

which can be rewritten into

$$\begin{aligned} \min \quad & \frac{1}{2}\begin{pmatrix} x^+ \\ x^- \end{pmatrix}^\top \begin{pmatrix} A^\top A & -A^\top A \\ -A^\top A & A^\top A \end{pmatrix}\begin{pmatrix} x^+ \\ x^- \end{pmatrix} + \begin{pmatrix} \mu\mathbb{1} - A^\top b \\ \mu\mathbb{1} + A^\top b \end{pmatrix}^\top \begin{pmatrix} x^+ \\ x^- \end{pmatrix} + \frac{1}{2}b^\top b \\ s.t. \quad & x^+, x^- \geq 0. \end{aligned} \tag{2.2}$$

The problem (2.1) can be solved by mosek and gurobi.

We plot the exact solution and solutions from cvx mosek, cvx gurobi, mosek and gurobi.

3 Various algorithms

3.1 Projection gradient method by reformulating the primal problem as a quadratic program with box constraints

According the the problem (2.1), define

$$f_\mu(x^+, x^-) = \frac{1}{2}(x^+ - x^-)^\top A^\top A(x^+ - x^-) + (\mu\mathbb{1} - A^\top b)^\top x^\top + (\mu\mathbb{1} + A^\top b)^\top x^- + \frac{1}{2}b^\top b \tag{3.1}$$

We aim to minimize $f_\mu(x^+, x^-)$ such that $x^+, x^- \geq 0$ by the projection gradient method. We have $\nabla_{x^+} f_\mu(x^+, x^-) = A^\top A(x^+ - x^-) + \mu\mathbb{1} - A^\top b$, $\nabla_{x^-} f_\mu(x^+, x^-) = A^\top A(x^- - x^+) + \mu\mathbb{1} + A^\top b$.

The projection gradient method can be summarized in Algorithm 1, where $\sigma(x) = \max\{x, 0\}$ is the projection operator.

Algorithm 1 Projection gradient method with continuation method

Input: initial value x_0, step size α, continuation parameter γ, N, maximum iteration number for each stage M.
for $i = 1, \cdots, N$ **do**
  $\mu_i = \gamma^{N-i}\mu$.
  for $j = 1, \cdots, M$ **do**
    $x^+ \leftarrow \sigma(x^+ - \alpha\nabla_{x^+} f_{\mu_i}(x^+, x^-)), x^- \leftarrow \sigma(x^- - \alpha\nabla_{x^-} f_{\mu_i}(x^+, x^-))$
  end for
end for
Output: $x = x^+ - x^-$.

3.2 Subgradient method for the primal problem

The subgradient of f_μ is $\partial f_\mu(x) = A^\top(Ax - b) + \mu \cdot \mathrm{sign}(x)$. The subgradient method can be summarized in Algorithm 2.

Algorithm 2 Subgradient method for the primal problem with continuation method

Input: initial value x_0, step size α, continuation parameter γ, N, maximum iteration number for each stage M.
for $i = 1, \cdots, N$ **do**
  $\mu_i = \gamma^{N-i}\mu$.
  for $j = 1, \cdots, M$ **do**
    $x \leftarrow x - \alpha\partial f_\mu(x)$.
  end for
end for
Output: x.

3.3 Gradient method for the smoothed primal problem

We consider the Huber penalty approximation of $\|x\|_1$:

$$\begin{aligned} h_\lambda(x) &= \sum_{i=1}^{n} h_\lambda^l(x) \\ \text{where} \quad h_\lambda^l(x) &= \begin{cases} x_l^2/(2\lambda), & |x_l| \le \lambda \\ |x_l| - \lambda/2, & |x_l| > \lambda \end{cases} \end{aligned} \tag{3.2}$$

We choose an additional parameter β for the decay of λ. The define $f_{i,j}(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu_i h_{\lambda_j}(x)$. The gradient can be computed as

$$\nabla f_{i,j}(x)_k = \begin{cases} (A^\top Ax - A^\top b)_k + \mu_i x_k/\lambda_j, & |x_k| \le \lambda_j, \\ (A^\top Ax - A^\top b)_k + \mu_i \mathrm{sign}(x_k), & |x_k| > \lambda_j \end{cases} \tag{3.3}$$

In $(k+1)$-th iteration, if $k = 0$, we use the initial step size α. Otherwise, we use the BB step size α_k:

$$\alpha_k = \frac{(x_k - x_{k-1})^\top(x_k - x_{k-1})}{(x_k - x_{k-1})^\top(\nabla f_{i,j}(x_k) - \nabla f_{i,j}(x_{k-1}))} \tag{3.4}$$

Then we can update x_{k+1} by $x_{k+1} = x_k - \alpha_k \nabla f_{i,j}(x_k)$.

Similarly,we use the continuation strategy. We have three parameters γ, M_1, M_2 for continuation, and set $\mu_0 = \mu_{\max} = \max\{\gamma\|A^\top b\|_\infty, \mu\}$. While $\mu_i > \mu$ or $\lambda_j > \lambda$, we update μ_{i+1}, λ_{i+1} by

$$\mu_{i+1} = \max\{\mu, \gamma\min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}, \quad \lambda_{j+1} = \max\{\beta\lambda_j, \lambda\} \tag{3.5}$$

Algorithm 3 Gradient method for smoothed primal problem with continuation strategy

Input: initial value x_0, step size α, continuation parameter γ, M_1, M_2, λ decay parameter β.
$\mu_0 = \mu_{\max} = \max\{\gamma \|A^\top b\|_\infty, \mu\}, \alpha_0 = \alpha, k = 0.$
while $\mu_i > \mu$ or $\lambda_j > \lambda$ **do**
  for $l = 1, 2, \cdots, M_1$ **do**
    Calculate BB step size α_k, and update x_{k+1}.
  end for
  $\mu_{i+1} = \max\{\mu, \gamma \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}, \lambda_{j+1} = \max\{\beta\lambda_j, \lambda\}, i = i+1, j = j+1.$
  Set $x_0 := x_k$ and $k = 0$. Update $\alpha_k = \min\{\alpha, \lambda_j\}$.
end while
for $l = 1, 2, \cdots, M_2$ **do**
  Calculate BB step size α_k, and update x_{k+1}.
end for

3.4 Fast (Nesterov/accelerated) gradient method for the smoothed primal problem

We still apply the continuation strategy with only a slight modification of the Algorithm 3. Specifically, we set $x_{-1} = x_0$. In $(k+1)$-th iteration, we update x_{k+1} by

$$\begin{cases} y & = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) \\ x_{k+1} & = y - \alpha_k \nabla f_{i,j}(x_k) \end{cases} \tag{3.6}$$

Algorithm 4 Fast gradient method for smoothed primal problem with continuation strategy

Input: initial value x_0, step size α, continuation parameter γ, M_1, M_2, λ decay parameter β.
$\mu_0 = \mu_{\max} = \max\{\gamma \|A^\top b\|_\infty, \mu\}, \alpha_0 = \alpha, k = 0.$
while $\mu_i > \mu$ or $\lambda_j > \lambda$ **do**
  for $l = 1, 2, \cdots, M_1$ **do**
    Update x_{k+1} by (3.6), $\alpha_{k+1} = \alpha_k, k = k+1$.
  end for
  $\mu_{i+1} = \max\{\mu, \gamma \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}, \lambda_{j+1} = \max\{\beta\lambda_j, \lambda\}, i = i+1, j = j+1.$
  Set $x_{=1} = x_0 := x_k$ and $k = 0$. Update $\alpha_k = \min\{\alpha, \lambda_j\}$.
end while
for $l = 1, 2, \cdots, M_2$ **do**
  Update x_{k+1} by (3.6), $\alpha_{k+1} = \alpha_k, k = k+1$.
end for

3.5 Proximal gradient method for the primal problem

Define the proximal operator $\text{prox}_{\alpha_k \mu h}(x) = \arg\min_z \frac{1}{2}\|z - x\|_2^2 + \alpha_k \mu h(z)$. When $h(x) = \|x\|_1$, the proximal operaator can be computed explicitly as $\text{prox}_{\alpha_k \mu h}(x) = S_{\alpha_k \mu}(x)$, where S is the soft thresholding operator. Define $f_i = g + \mu_i h$, in $(k+1)$-th iteration, we use the BB step size

$$\alpha_k = \frac{(x_k - x_{k-1})^\top (x_k - x_{k-1})}{(x_k - x_{k-1})^\top (\nabla g(x_k) - \nabla g(x_{k-1}))} \tag{3.7}$$

Then, we update x_{k+1} by

$$x_{k+1} = S_{\alpha_k \mu_i}(x_k - \alpha_k \nabla g(x_k)) \tag{3.8}$$

3.6 Fast proximal gradient method for the primal problem

In this part, we update x_{k+1} by

$$\begin{cases} y_k & = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}) \\ x_{k+1} & = S_{\alpha_k \mu_i}(y_k - \alpha_k \nabla g(y_k)) \end{cases} \tag{3.9}$$

Algorithm 5 Proximal gradient method with continuation strategy

Input: initial value x_0, step size α, continuation parameter $\gamma, \varepsilon_1, \varepsilon_2$.
$\mu_0 = \mu_{\max} = \max\{\gamma \|A^\top b\|_\infty, \mu\}, \alpha_0 = \alpha, i = k = 0$.
Update x_{k+1} by (3.8), $k = k + 1$.
while $\mu_i > \mu$ **do**
  for $k = 1, 2, \cdots, M_1$ **do**
    Calculate BB step size s_k by (3.7), update x_{k+1} by (3.8).
  end for
  $\mu_{i+1} = \max\{\mu, \gamma \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}, i = i + 1$.
  Set $\alpha_k = \alpha$, update x_{k+1} by (3.8), $k = k + 1$.
end while
for $k = 1, 2, \cdots, M_2$ **do**
  Calculate BB step size s_k by (3.7), update x_{k+1} by (3.8).
end for

Algorithm 6 Fast proximal gradient method with continuation strategy

Input: initial value x_0, step size α, continuation parameter $\gamma, \varepsilon_1, \varepsilon_2$.
$\mu_0 = \mu_{\max} = \max\{\gamma \|A^\top b\|_\infty, \mu\}, \alpha_0 = \alpha, i = k = 0$.
Update x_{k+1} by (3.9), $k = k + 1$.
while $\mu_i > \mu$ **do**
  for $k = 1, 2, \cdots, M_1$ **do**
    Calculate BB step size s_k by (3.7), update x_{k+1} by (3.9).
  end for
  $\mu_{i+1} = \max\{\mu, \gamma \min\{\|\nabla g(x_k)\|_\infty, \mu_i\}\}, i = i + 1$.
  Set $\alpha_k = \alpha$, update x_{k+1} by (3.9), $k = k + 1$.
end while
for $k = 1, 2, \cdots, M_2$ **do**
  Calculate BB step size s_k by (3.7), update x_{k+1} by (3.9).
end for

3.7 Augmented Lagrangian method for the dual problem

The original problem (1.1) is equivalent to the following problem:

$$\min_x \quad \frac{1}{2}\|y\|_2^2 + \mu\|x\|_1 \quad \text{s.t.} \quad Ax - b = y \tag{3.10}$$

The corresponding Lagrangian is

$$L(x, y, z) = \frac{1}{2}\|y\|_2^2 + \mu\|x\|_1 + z^\top(Ax - b - y) \tag{3.11}$$

where $z \in \mathbb{R}^m$. By minimizing L, we have

$$\begin{aligned} \min_{x,y} L(x, y, z) &= -b^\top z + \min_y(\frac{1}{2}\|y\|_2^2 - z^\top y) + \min_x(\mu h(x) + (A^\top z)^\top x) \\ &= -b^\top z - g_0^\star(z) - \mu h^\star(A^\top z/\mu) \end{aligned} \tag{3.12}$$

where the $g_0^\star$ and $h^\star$ are the conjugate of the function $g_0 = \frac{1}{2}\|\cdot\|_2^2$ and h, which can be directly computed by $g_0^\star(z) = \frac{1}{2}\|z\|^2$, $h^\star(z) = \begin{cases} 0, & \|z\|_\infty \le 1 \\ +\infty, & \|z\|_\infty > 1 \end{cases}$

Hence the dual problem for problem (1.1) is

$$\min \frac{1}{2}\|z\|_2^2 + b^\top z, \quad \text{s.t.} \quad A^\top z = w, \quad \|w\|_\infty \le \mu. \tag{3.13}$$

whose augmented Lagrangian is

$$L_a(z, w, \lambda) = \frac{1}{2}\|z\|_2^2 + b^\top z + \lambda^\top(A^\top z - w) + \frac{a}{2}\|A^\top z - w\|_2^2. \tag{3.14}$$

If we set $z^0 = 0, w^0, \lambda^0 = 0$. Given (z^k, w^k, λ^k), the relationship between w^{k+1} and z^{k+1}

$$w^{k+1} = \lambda^k/a + A^\top z^{k+1} - S_\mu(\lambda^k/a + A^\top z^{k+1}). \tag{3.15}$$

where the soft thresholding function S_μ is defined as

$$S_\mu(w) = \mathrm{sign}(w) \cdot (|w| - \mu)^+ \tag{3.16}$$

Then, we have the following problem:

$$\arg\min_z \frac{1}{2}\|z\|_2^2 + b^\top z + \frac{a}{2}\|S_\mu(\lambda^k/t + A^\top z)\|_2^2 \tag{3.17}$$

We consider to use the Newton's method to solve the minimization (3.17). We define $z^{k,0} = z^k$, the update can be written as

$$\begin{aligned} z^{k,j+1} &= z^{k,j} - H(z^{k,j})^{-1} d(z^{k,j}) \\ &= z^{k,j} - (z^{k,j} + b + a \sum_{|v_i^{k,j}|>\mu} A_i S_\mu(v^{k,j})_i)^{-1} (I + a \sum_{|v_i^{k,j}|>\mu} A_i A_i^\top) \end{aligned} \tag{3.18}$$

where $v^{k,j} = \lambda^k/a + A^\top z^{k,j}$. We perform the update until $\|d(z^{k,j})\|_2 / \|d(z^{k,0})\|_2 \leq \epsilon_3$, assuming we terminate the iteration at the M_2-th step.

Since the computational cost of solving $H(z^{k,j})^{-1} d(z^{k,j})$ is large when $H(z^{k,j})$ varies, we approximate $H(z^{k,j}) \approx I + aAA^\top = LDL^\top$ in advance. Empirically, we find approximate $d(z^{k,j}) \approx z^{k,j} + b + aAS_\mu(v^{k,j})$ does not impair the performance and improve the efficiency.

In all, we can update $(z^{k+1}, w^{k+1}, \lambda^{k+1})$:

$$\begin{cases} z^{k+1} = z^{k,M_3}. \\ w^{k+1} = \lambda^k/a + A^\top z^{k+1} - S_\mu(\lambda^k/a + A^\top z^{k+1}) \\ \lambda^{k+1} = \lambda^k + a(A^\top z^{k+1} - w^{k+1}) \end{cases} \tag{3.19}$$

Algorithm 7 Augmented Lagrangian method for the dual problem with continuation strategy.

Input: Augmented Lagragian parameter a, continuation parameter γ, M_1, M_2, Newton's method parameter M_3. Calculate $\mu_0 = \max\{\gamma\|A^\top b\|_\infty, \mu\}$. Initialize variables $i = k = 0, z^0 = 0, \lambda^0 = 0$.
while $\mu_i > \mu$ **do**
  for $k = 1, 2, \cdots, M_1$ **do**
    Update $(z^{k+1}, w^{k+1}, \lambda^{k+1})$ by (3.19), $k = k + 1$.
  end for
  $\mu_{i+1} = \max\{\mu, \gamma\mu_i\}, i = i + 1, z^0 = z^k, \lambda^0 = \lambda^k, k = 0$.
end while
for $k = 1, 2, \cdots, M_2$ **do**
  Update $(z^{k+1}, w^{k+1}, \lambda^{k+1})$ by (3.19), $k = k + 1$.
end for
$x = -\lambda^k$.

3.8 Alternating direction method of multipliers for the dual problem

Similarity we obtain the augmented Lagrangian (3.14), while we minimize this Lagrangian with alternating direction strategy. First we minimize $L_a(z^k, w, \lambda^k)$ w.r.t. w, we have $W^{k+1} = \lambda^k/a + A^\top z^k - S_\mu(\lambda^k/a + A^\top z^k)$. Then we minimize $L_a(w^{k+1}, z, \lambda^k)$ w.r.t. z. Therefore we can update $(z^{k+1}, w^{k+1}, \lambda^{k+1})$:

$$\begin{cases} w^{k+1} = \lambda^k/a + A^\top z^k - S_\mu(\lambda^k/a + A^\top z^k) \\ z^{k+1} = (I + aAA^\top)^{-1}(-b - A\lambda^k + aAw^{k+1}) \\ \lambda^{k+1} = \lambda^k + a(A^\top z^{k+1} - w^{k+1}) \end{cases} \tag{3.20}$$

Algorithm 8 ADMM for the dual problem with continuation strategy

Input: Augmented Lagragian parameter a, continuation parameter γ, M_1, M_2. Calculate $\mu_0 = \max\{\gamma\|A^\top b\|_\infty, \mu\}$. Initialize variables $i = k = 0, z^0 = 0, \lambda^0 = 0$.
while $\mu_i > \mu$ **do**
  for $k = 1, 2, \cdots, M_1$ **do**
    Update $(z^{k+1}, w^{k+1}, \lambda^{k+1})$ by (3.20).
  end for
  $\mu_{i+1} = \max\{\mu, \gamma\mu_i\}, i = i + 1, z^0 = z^k, \lambda^0 = \lambda^k, k = 0.$
end while
for $k = 1, 2, \cdots, M_2$ **do**
  Update $(z^{k+1}, w^{k+1}, \lambda^{k+1})$ by (3.20).
end for
$x = -\lambda^k$.

3.9 Alternating direction method of multipliers with linearization for the primal problem

The primal problem can be reformulated as

$$\min \frac{1}{2}\|Ax - b\|_2^2 + \mu\|y\|_1 \quad \text{s.t. } x = y \tag{3.21}$$

The augmented Lagrangian is

$$L_a^p(x, y, z) = \frac{1}{2}\|Ax - b\|_2^2 + \mu\|y\|_1 + z^\top(x - y) + \frac{a}{2}\|x - y\|_2^2. \tag{3.22}$$

We first update x^{k+1} by direct minimization $x^{k+1} = \arg\min_x L_a(x^k, y^k, z^k) = (A^\top A + aI)^{-1}(A^\top b + ay^k - z^k)$; then we update $y^{k+1} = \arg\min_y L_a(x^{k+1}, y, z^k) = S_{\frac{\mu}{a}}(x^{k+1} + \frac{z^k}{t})$. The update can be summarized as

$$\begin{cases} x^{k+1} = (A^\top A + aI)^{-1}(A^\top b + ay^k - z^k) \\ y^{k+1} = S_{\frac{\mu}{a}}(x^{k+1} + \frac{z^k}{a}) \\ z^{k+1} = z^k + a(x^{k+1} - y^{k+1}) \end{cases} \tag{3.23}$$

Algorithm 9 ADMM with linearization for the primal problem with continuation strategy

Input: Augmented Lagragian parameter a, continuation parameter $\gamma, \varepsilon_1, \varepsilon_2$. Calculate $\mu_0 = \max\{\gamma\|A^\top b\|_\infty, \mu\}$. Initialize variables $i = k = 0, x^0 = y^0 = x_0, z^0 = 0$.
while $\mu_i > \mu$ **do**
  for $k = 1, 2, \cdots, M_1$ **do**
    Update $(x^{k+1}, y^{k+1}, z^{k+1})$ by (3.23), $k = k + 1$.
  end for
  $\mu_{i+1} = \max\{\mu, \gamma\mu_i\}, i = i + 1, x^0 = z^k, y^0 = y^k, z^0 = z^k, k = 0.$
end while
for $k = 1, 2, \cdots, M_2$ **do**
  Update $(x^{k+1}, y^{k+1}, z^{k+1})$ by (3.23), $k = k + 1$.
end for
$x = x^k$.

References