CONVEX OPTIMIZATION HW

SPRING 2025

1. LASSO Problem numerical optimization

   Solution. *See jupyter notebook file.*

2. Write as a linear program.
   (a) $\min \|A\mathbf{x} - \mathbf{b}\|_1$.

   Solution. Tthe 1-norm can be expressed as

$$\|A\mathbf{x} - \mathbf{b}\|_1 = \sum_{i=1}^{m} |(A\mathbf{x} - \mathbf{b})_i|.$$

   To handle the absolute values, we introduce nonnegative auxiliary variables t_i for $i = 1, 2, \ldots, m$, and require

$$t_i \geq (A\mathbf{x} - \mathbf{b})_i \quad \text{and} \quad t_i \geq -(A\mathbf{x} - \mathbf{b})_i,$$

   with $t_i \geq 0$. This allows us to reformulate the original problem as

$$\min_{\mathbf{x},t} \sum_{i=1}^{m} t_i,$$

   subject to

$$A\mathbf{x} - \mathbf{b} \leq t,$$
$$-A\mathbf{x} + \mathbf{b} \leq t,$$
$$t \geq 0,$$

   where t denotes the vector with components t_i.

   (b) $\min \|\mathbf{x}\|_1$ s.t $\|A\mathbf{x} - \mathbf{b}\|_\infty \leq 1$.

   Solution. Again, note that the 1-norm of $\mathbf{x}$ is given by

$$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |x_j|.$$

   To handle the absolute values, we introduce nonnegative auxiliary variables $u_j \geq 0$ for $j = 1, 2, \ldots, n$, and impose

$$u_j \geq x_j \quad \text{and} \quad u_j \geq -x_j.$$

   Next, the constraint $\|A\mathbf{x} - \mathbf{b}\|_\infty \leq 1$ means that the maximum absolute value among the entries of $A\mathbf{x} - \mathbf{b}$ is at most 1. In other words, for each row i, we have

$$|(A\mathbf{x} - \mathbf{b})_i| \leq 1.$$

   This can be split into the two linear inequalities

$$A\mathbf{x} - \mathbf{b} \leq \mathbf{1}, \quad \text{and} \quad -A\mathbf{x} + \mathbf{b} \leq \mathbf{1},$$

where $\mathbf{1}$ is the vector of ones. Thus, the final linear programming formulation is given by

$$\begin{aligned}
\min_{\mathbf{u},\mathbf{x}} \quad & \sum_{j=1}^{n} u_j \\
\text{subject to} \quad & u_j \geq x_j, \quad j = 1, \ldots, n, \\
& u_j \geq -x_j, \quad j = 1, \ldots, n, \\
& A\mathbf{x} - \mathbf{b} \leq \mathbf{1}, \\
& -A\mathbf{x} + \mathbf{b} \leq \mathbf{1}, \\
& u_j \geq 0, \quad j = 1, \ldots, n.
\end{aligned}$$

3. Lagrange Duality

Solution. **Feasible Set:** The inequality

$$(x-2)(x-4) \leq 0$$

has roots at $x = 2$ and $x = 4$, and it holds for

$$x \in [2, 4].$$

Optimal Solution: Since $f(x) = x^2 + 1$ is strictly convex, and its stationary point is not in the feasibility set, so its minimum in the feasible set is attained at one of the boundary points. Evaluating,

$$f(2) = 2^2 + 1 = 5, \quad f(4) = 4^2 + 1 = 17.$$

Thus, the optimal solution is $x = 2$ with the optimal value $f(2) = 5$. Introduce the Lagrange multiplier $\lambda \geq 0$ for the constraint. The Lagrangian is

$$\begin{aligned}
L(x, \lambda) &= x^2 + 1 + \lambda(x-2)(x-4) \\
&= x^2 + 1 + \lambda(x^2 - 6x + 8) \\
&= (1+\lambda)x^2 - 6\lambda x + (1 + 8\lambda).
\end{aligned}$$

The dual function is defined as

$$g(\lambda) = \inf_{x \in \mathbb{R}} L(x, \lambda).$$

Since $L(x, \lambda)$ is quadratic in x with $1 + \lambda > 0$ (for $\lambda \geq 0$), its minimum occurs at

$$\frac{\partial L}{\partial x} = 2(1+\lambda)x - 6\lambda = 0 \quad \Longrightarrow \quad x^* = \frac{3\lambda}{1+\lambda}.$$

Substitute x^* into $L(x, \lambda)$:

$$\begin{aligned}
g(\lambda) &= L\left(\frac{3\lambda}{1+\lambda}, \lambda\right) \\
&= (1+\lambda)\left(\frac{3\lambda}{1+\lambda}\right)^2 - 6\lambda\left(\frac{3\lambda}{1+\lambda}\right) + (1+8\lambda) \\
&= \frac{9\lambda^2}{1+\lambda} - \frac{18\lambda^2}{1+\lambda} + (1+8\lambda) \\
&= 1 + 8\lambda - \frac{9\lambda^2}{1+\lambda} = \frac{1 + 9\lambda - \lambda^2}{1+\lambda}, \quad \lambda \geq 0.
\end{aligned}$$

Dual Problem: The dual problem is

$$\max_{\lambda \geq 0} g(\lambda).$$

The dual function $g(\lambda)$ is the pointwise infimum of a family of functions affine in λ. Therefore, g. To maximize $g(\lambda)$ for $\lambda \geq 0$, we differentiate using the quotient rule. Define

$$N(\lambda) = 1 + 9\lambda - \lambda^2, \quad D(\lambda) = 1 + \lambda.$$

Then

$$g'(\lambda) = \frac{N'(\lambda)D(\lambda) - N(\lambda)D'(\lambda)}{(D(\lambda))^2},$$

with

$$N'(\lambda) = 9 - 2\lambda \quad \text{and} \quad D'(\lambda) = 1.$$

Thus,

$$\begin{aligned} g'(\lambda) &= \frac{(9-2\lambda)(1+\lambda) - (1 + 9\lambda - \lambda^2)}{(1+\lambda)^2} \\ &= \frac{[9 + 7\lambda - 2\lambda^2] - (1 + 9\lambda - \lambda^2)}{(1+\lambda)^2} \\ &= \frac{8 - 2\lambda - \lambda^2}{(1+\lambda)^2}. \end{aligned}$$

Setting the numerator equal to zero,

$$8 - 2\lambda - \lambda^2 = 0 \quad \Longrightarrow \quad \lambda^2 + 2\lambda - 8 = 0.$$

The solutions are

$$\lambda = \frac{-2 \pm \sqrt{4 + 32}}{2} = \frac{-2 \pm 6}{2},$$

giving $\lambda = 2$ and $\lambda = -4$. Since $\lambda \geq 0$ we take $\lambda^* = 2$. Evaluating the dual function at $\lambda = 2$,

$$g(2) = \frac{1 + 9 \cdot 2 - 2^2}{1 + 2} = \frac{1 + 18 - 4}{3} = \frac{15}{3} = 5.$$

Strong Duality: The optimal value of the primal problem is $f(2) = 2^2 + 1 = 5$. Since the dual optimal value is also 5, strong duality holds.

For the plots, see the jupyter notebook file.

4. Duality of Boolean LP
   (a) Lagrange Duality

   Solution. The Lagrangian is

$$\begin{aligned} L(x, \mu, \nu) &= c^\top x + \mu^\top (Ax - b) - \nu^\top x + x^\top \mathrm{diag}(\nu) x \\ &= x^\top \mathrm{diag}(\nu) x + (c + A^\top \mu - \nu)^\top x - b^\top \mu. \end{aligned}$$

   Setting $\nabla_x f = 0$ gives

$$2\,\mathrm{diag}(\nu) x^* + (c + A^\top \mu - \nu) = 0 \Longrightarrow x^* = -\tfrac{1}{2}\,\mathrm{diag}(\nu)^{-1}(c + A^\top \mu - \nu).$$

   Substituting back:

$$\begin{aligned} x^{*\top}\,\mathrm{diag}(\nu) x^* &= \tfrac{1}{4}(c + A^\top \mu - \nu)^\top\,\mathrm{diag}(\nu)^{-1}(c + A^\top \mu - \nu), \\ (c + A^\top \mu - \nu)^\top x^* &= -\tfrac{1}{2}(c + A^\top \mu - \nu)^\top\,\mathrm{diag}(\nu)^{-1}(c + A^\top \mu - \nu), \end{aligned}$$

   so their sum is

$$x^{*\top}\,\mathrm{diag}(\nu) x^* + (c + A^\top \mu - \nu)^\top x^* = -\tfrac{1}{4}(c + A^\top \mu - \nu)^\top\,\mathrm{diag}(\nu)^{-1}(c + A^\top \mu - \nu).$$

Hence the dual function is

$$g(\mu,\nu) = \inf_x L(x,\mu,\nu) = -b^\top \mu - \tfrac{1}{4}(c + A^\top \mu - \nu)^\top \operatorname{diag}(\nu)^{-1}(c + A^\top \mu - \nu),$$

or, in component form,

$$g(\mu,\nu) = -b^\top \mu - \tfrac{1}{4}\sum_{i=1}^n \frac{\left(c_i + a_i^\top \mu - \nu_i\right)^2}{\nu_i}, \quad \nu \succeq 0.$$

Thus minimizing over x gives the dual function

$$g(\mu,\nu) = \begin{cases} -b^\top \mu - (1/4)\sum_{i=1}^n (c_i + a_i^\top \mu - \nu_i)^2/\nu_i & \nu \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

where a_i is the ith column of A. The resulting dual problem is

$$\begin{aligned} &\text{maximize} \quad -b^\top \mu - \frac{1}{4}\sum_{i=1}^n \frac{(c_i + a_i^\top \mu - \nu_i)^2}{\nu_i} \\ &\text{subject to} \quad \nu \succeq 0, \quad \mu \succeq 0. \end{aligned}$$

For each i, define

$$h_i(\nu_i) = -\frac{(c_i + a_i^\top \mu - \nu_i)^2}{\nu_i}, \quad \nu_i \geq 0.$$

We maximize h_i by setting

$$0 = \frac{d}{d\nu_i}\left[-\tfrac{(c_i + a_i^\top \mu - \nu_i)^2}{\nu_i}\right] = \frac{(c_i + a_i^\top \mu)^2}{\nu_i^2} - 1 \quad \Longrightarrow \quad \nu_i^* = \left|c_i + a_i^\top \mu\right|.$$

Substituting back,

$$\sup_{\nu_i \geq 0} h_i(\nu_i) = \begin{cases} 4(c_i + a_i^\top \mu), & c_i + a_i^\top \mu \leq 0, \\ 0, & c_i + a_i^\top \mu \geq 0, \end{cases} = \min\{0, 4(c_i + a_i^\top \mu)\}.$$

Therefore the ν-optimized dual becomes

$$\begin{aligned} \sup_{\nu \succeq 0} g(\mu,\nu) &= -b^\top \mu - \tfrac{1}{4}\sum_{i=1}^n \sup_{\nu_i \geq 0}\left[-\tfrac{(c_i + a_i^\top \mu - \nu_i)^2}{\nu_i}\right] \\ &= -b^\top \mu + \sum_{i=1}^n \min\{0, c_i + a_i^\top \mu\}. \end{aligned}$$

So the final dual problem is

$$\begin{aligned} &\max_\mu -b^\top \mu + \sum_{i=1}^n \min\{0, c_i + a_i^\top \mu\}, \\ &\text{s.t} \quad \mu \succeq 0. \end{aligned}$$

(b) LP relaxation

Solution. The Lagrangian function of the LP relaxation is

$$\begin{aligned} L(x,u,v,w) &= c^\top x + u^\top (Ax - b) - v^\top x + w^\top (x - 1) \\ &= (c + A^\top u - v + w)^\top x - b^\top u - \mathbf{1}^\top w, \end{aligned}$$

where v and w are Lagrange multipliers. Then minizing over x by first order condition gives the dual objective function

$$g(u,v,w) = \begin{cases} -b^\top u - \mathbf{1}^\top w & A^\top u - v + w + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

The dual problem is therefore

$$\begin{aligned} \text{maximize} \quad & -b^\top u - \mathbf{1}^\top w \\ \text{subject to} \quad & A^\top u - v + w + c = 0 \\ & u \succeq 0,\ v \succeq 0,\ w \succeq 0, \end{aligned}$$

which is equivalent to the Lagrange relaxation problem derived above by marginally optimizing v and w first.

5. Calculate subgradient
   (a) $f(\mathbf{x}) = \max_i |\mathbf{a}_i \mathbf{x} + b_i|$

   Solution. Let

$$r_i = \mathbf{a}_i^\top \mathbf{x} + b_i, \quad i = 1, \dots, m,$$

   and let

$$f(\mathbf{x}) = \max_i |r_i| = \alpha.$$

   Define the active set $I \subseteq \{1, \dots, m\}$ by

$$I = \{i | |r_i| = \alpha\}.$$

   For each i in the active set I, the subgradient for the term $|\mathbf{a}_i^\top \mathbf{x} + b_i|$ can be written as

$$\mathbf{g}_i = \operatorname{sign}(r_i)\mathbf{a}_i = \begin{cases} +\mathbf{a}_i, & r_i > 0, \\ -\mathbf{a}_i, & r_i < 0, \\ \text{any vector in } [-1, +1]\mathbf{a}_i, & r_i = 0. \end{cases}$$

   Because $f(\mathbf{x})$ is the maximum of these absolute-value terms, any subgradient $\mathbf{g}$ of f at $\mathbf{x}$ must lie in the *convex hull* of the $\mathbf{g}_i$'s coming from the active indices:

$$\mathbf{g} \in \operatorname{conv}\Big\{\operatorname{sign}(r_i)\mathbf{a}_i : i \in I\Big\}.$$

   In other words, pick any convex combination of the vectors $\operatorname{sign}(r_i)\mathbf{a}_i$ for i in I.

   (b) $f(\mathbf{x}) = \mathbf{x}_{[1]} + \cdots + \mathbf{x}_{[k]}$, where $\mathbf{x}_{[1]} \geq \mathbf{x}_{[2]} \geq \cdots \geq \mathbf{x}_{[n]}$ denote the components of $\mathbf{x}$ in descending order.

   Solution. Let S be the set of indices corresponding to the k largest entries of $\mathbf{x}$. That is,

$$S = \big\{i : x_i \text{ is among the top } k \text{ values of } \mathbf{x}\big\}.$$

   Suppose $\mathbf{x}_{[k]} > \mathbf{x}_{[k+1]}$. Then S has exactly k distinct elements, and a subgradient is the vector

$$\mathbf{g} \quad \text{where} \quad g_i = \begin{cases} 1, & i \in S, \\ 0, & i \notin S. \end{cases}$$

   If there is a tie for the k-th largest value (i.e., $\mathbf{x}_{[k]} = \mathbf{x}_{[k+1]}$), we let:

$$\ell = \#\{i : x_i > \mathbf{x}_{[k]}\}, \quad T = \{i : x_i = \mathbf{x}_{[k]}\}.$$

   Then we must assign $g_i = 1$ to indices strictly greater than $\mathbf{x}_{[k]}$, and $g_i = 0$ for indices strictly less. For i in T, assign g_i in $[0, 1]$ so that

$$\sum_{i \in T} g_i = k - \ell.$$

   Hence the subgradient belongs to the convex set of all such assignments.

   (c) $f(\mathbf{x}) = \inf_{A\mathbf{y} \preceq \mathbf{b}} \|\mathbf{x} - \mathbf{y}\|_2^2$.

Solution. This is the squared distance from $\mathbf{x}$ to the polyhedron $C = \{\mathbf{y} : A\mathbf{y} \preceq \mathbf{b}\}$. Let

$$f(\mathbf{x}) = \|\mathbf{x} - \Pi_C(\mathbf{x})\|_2^2$$

where $\Pi_C(\mathbf{x})$ denotes the Euclidean projection of $\mathbf{x}$ onto C:

$$\Pi_C(\mathbf{x}) = \arg\min_{\mathbf{y}\in C} \|\mathbf{x} - \mathbf{y}\|_2^2.$$

If $\mathbf{x} \in C$. Then $f(\mathbf{x}) = 0$ and $\mathbf{x}$ itself is a closest point in C. In a neighborhood inside C, f stays 0, so the subgradient at $\mathbf{x}$ is $\mathbf{0}$. If $\mathbf{x} \notin C$. Let $\mathbf{y}^* \in \Pi_C(\mathbf{x})$. Then

$$f(\mathbf{x}) = \|\mathbf{x} - \mathbf{y}^*\|_2^2.$$

The gradient (w.r.t. $\mathbf{x}$) of $\|\mathbf{x} - \mathbf{y}^*\|_2^2$ is

$$\nabla_{\mathbf{x}}\|\mathbf{x} - \mathbf{y}^*\|_2^2 = 2(\mathbf{x} - \mathbf{y}^*).$$

Thus a subgradient of f at $\mathbf{x}$ is

$$\mathbf{g} = 2(\mathbf{x} - \mathbf{y}^*),$$

for the $\mathbf{y}^*$ that is the projection $\Pi_C(\mathbf{x})$.

(d) $f(\mathbf{x}) = \sup_{A\mathbf{y}\preceq\mathbf{b}} \mathbf{y}^\top\mathbf{x}$.

Solution. Since $f(\mathbf{x}) = \sup_{\mathbf{y}\in P} \mathbf{y}^\top\mathbf{x}$ is just the support function of $P := \{A\mathbf{y} \preceq \mathbf{b}\}$, we immediately conclude:

$$\partial f(\mathbf{x}) = \operatorname{conv}\big(\{\mathbf{y}^* \in P : \mathbf{y}^{*\top}\mathbf{x} = f(\mathbf{x})\}\big).$$

Let $\mathbf{y}^*$ be any optimal solution of $\max\limits_{\mathbf{y}\in P} \mathbf{y}^\top\mathbf{x}$. If there is a unique solution, the subgradient is just $\{\mathbf{y}^*\}$.

6. Subgradient of $\|x\|_2$.

Solution. When $x \neq 0$, f is differentiable at x. Recall that for $x \neq 0$, the gradient is

$$\nabla f(x) = \frac{x}{\|x\|_2}.$$

For a convex function that is differentiable at $x \neq 0$, the subdifferential coincides with the set containing its gradient. Hence,

$$\partial f(x) = \left\{\frac{x}{\|x\|_2}\right\}.$$

When $x = 0$, the function f is not differentiable. We use the definition of a subgradient for a convex function: g is in the subdifferential $\partial f(0)$ if and only if

$$f(y) \geq f(0) + \langle g, y - 0\rangle \quad \text{for all } y.$$

Since $f(0) = 0$, we need

$$\|y\|_2 \geq \langle g, y\rangle \quad \text{for all } y.$$

By Cauchy-Schwarz inequality, for all g such that $\|g\|_2 \leq 1$ we obtain

$$\|y\|_2 \geq \|y\|_2\|g\|_2 \geq \langle g, y\rangle,$$

we therefore have

$$\partial f(0) = \{g : \|g\|_2 \leq 1\}.$$

7. Compute proximal operators.

(a) $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c, \quad$ where $A \in S_n^+$.

Solution. We wish to compute the proximal operator defined by

$$\mathrm{prox}_f(v) = \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{1}{2}\|x-v\|_2^2 \right\}.$$

We wish to solve

$$\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}x^\top Ax + b^\top x + c + \frac{1}{2}\|x-v\|_2^2 \right\}.$$

Define

$$\varphi(x) = \frac{1}{2}x^\top Ax + b^\top x + \frac{1}{2}\|x-v\|_2^2.$$

Taking the gradient with respect to x, we have

$$\nabla\varphi(x) = Ax + b + (x-v).$$

Setting the gradient equal to zero gives:

$$Ax + b + x - v = 0.$$

Solving for the optimal solution

$$\mathrm{prox}_f(v) = x = (A+I)^{-1}(v-b).$$

(b) $f(x) = -\sum_{i=1}^n \log x_i$ where $x \in \mathbb{R}^n_+$.

Solution. We want to compute

$$\mathrm{prox}_f(v) = \arg\min_{x\in\mathbb{R}^n_+} \Big\{ -\sum_{i=1}^n \log x_i + \tfrac{1}{2}\|x-v\|_2^2 \Big\}.$$

Since the objective is separable, we solve the scalar subproblem for each i:

$$x_i' = \arg\min_{x_i>0} \Big\{ -\log x_i + \tfrac{1}{2}(x_i - v_i)^2 \Big\}.$$

Taking the derivative with respect to x_i and setting it to zero:

$$-\tfrac{1}{x_i} + (x_i - v_i) = 0 \quad \Longrightarrow \quad x_i^2 - v_i x_i - 1 = 0.$$

Solving the quadratic equation for x_i yields

$$x_i = \frac{v_i \pm \sqrt{v_i^2+4}}{2}.$$

Since $x_i > 0$, we take

$$x_i' = \frac{v_i + \sqrt{v_i^2+4}}{2}.$$

Hence, the proximal operator of f evaluated at v is given coordinate-wise by

$$(\mathrm{prox}_f(v))_i = \frac{v_i + \sqrt{v_i^2+4}}{2}, \quad i = 1, \ldots, n.$$

(c) $f(x) = \|x\|_2$.

Solution. We wish to solve

$$\mathrm{prox}_f(v) = \arg\min_{x\in\mathbb{R}^n} \Big\{ \|x\|_2 + \tfrac{1}{2}\|x-v\|_2^2 \Big\}.$$

A key observation is that, because the Euclidean norm $\|\cdot\|_2$ is rotationally symmetric, the minimizer x must lie on the same ray as v. Thus we can write

$$x = tv \quad \text{for some scalar } t \geq 0.$$

Substituting $x = tv$ into the objective function yields

$$\|x\|_2 = t\|v\|_2, \quad \|x - v\|_2^2 = \|(t-1)v\|_2^2 = (t-1)^2\|v\|_2^2.$$

Hence the optimization problem reduces to

$$\min_{t\geq 0}\Big\{t\|v\|_2 + \tfrac{1}{2}(t-1)^2\|v\|_2^2\Big\}.$$

Let us define

$$\phi(t) = t\|v\|_2 + \tfrac{1}{2}(t-1)^2\|v\|_2^2.$$

Taking the derivative with respect to t:

$$\frac{d\phi}{dt} = \|v\|_2 + (t-1)\|v\|_2^2.$$

Setting this derivative to zero for an unconstrained minimizer:

$$\|v\|_2 + (t-1)\|v\|_2^2 = 0 \Longrightarrow t = 1 - \frac{1}{\|v\|_2}.$$

Because $t \geq 0$, we require

$$1 - \frac{1}{\|v\|_2} \geq 0 \quad \Longrightarrow \quad \|v\|_2 \geq 1.$$

Thus,

$$t = \begin{cases} 1 - \frac{1}{\|v\|_2}, & \text{if } \|v\|_2 \geq 1, \\ 0, & \text{if } \|v\|_2 < 1. \end{cases}$$

Then the result can be written as

$$\mathrm{prox}_{\|\cdot\|_2}(v) = \big(1 - \tfrac{1}{\|v\|_2}\big)_+ v.$$

(d) $f(x) = \|x\|_0$, where $\|x\|_0 = \mathrm{card}\{x_i : x_i \neq 0, i = 1, \ldots, n\}$.

Solution. We want

$$\mathrm{prox}_f(v) = \arg\min_{x\in\mathbb{R}^n}\Big\{\|x\|_0 + \tfrac{1}{2}\|x - v\|_2^2\Big\},$$

with $\|x\|_0 = \mathrm{card}\{i : x_i \neq 0\}$. Because the objective is separable, consider the scalar problem for each coordinate i:

$$\min_{x_i\in\mathbb{R}}\Big\{\mathbf{1}_{\{x_i\neq 0\}} + \tfrac{1}{2}(x_i - v_i)^2\Big\}.$$

Here $\mathbf{1}_{\{x_i\neq 0\}} = 0$ if $x_i = 0$, and 1 otherwise. Thus if $x_i = 0$, the cost is $\frac{1}{2}v_i^2$. If $x_i \neq 0$, the optimal choice is $x_i = v_i$ and the cost is 1. Hence we compare $\frac{1}{2}v_i^2$ with 1. If $\frac{1}{2}v_i^2 \leq 1$, it is cheaper to set $x_i = 0$. Otherwise, choose $x_i = v_i$. That is,

$$(\mathrm{prox}_f(v))_i = \begin{cases} 0, & |v_i| \leq \sqrt{2}, \\ v_i, & |v_i| > \sqrt{2}. \end{cases}$$

8. Proximal Operator of The Group Lasso Problem

Solution. We take the definition

$$\mathrm{prox}_{h,t}(x) = \arg\min_{z\in\mathbb{R}^p}\Big\{\frac{1}{2t}\big\|x - z\big\|_2^2 + h(z)\Big\},$$

where

$$h(z) = \lambda\sum_{j=1}^{J}\omega_j\|z_{(j)}\|_2,$$

and $z_{(j)}$ is the jth block of z. Since h is separable over the blocks, the minimization decouples:

$$\big[\mathrm{prox}_{h,t}(x)\big]_{(j)} = \arg\min_{z_{(j)}\in\mathbb{R}^{d_j}} \frac{1}{2t}\big\|x_{(j)} - z_{(j)}\big\|_2^2 + \lambda\omega_j\big\|z_{(j)}\big\|_2.$$

Define

$$\alpha_j = t\lambda\omega_j,$$

and note that multiplying the objective by the positive constant t does not change the minimizer. Equivalently we solve

$$z^* = \arg\min_{z\in\mathbb{R}^{d_j}} \underbrace{\frac{1}{2}\|z-x\|_2^2 + \alpha_j\|z\|_2}_{\Phi(z)}, \quad x = x_{(j)}.$$

(i) If $x = 0$, clearly $z^* = 0$.

(ii) Otherwise write

$$u = \frac{x}{\|x\|_2}, \quad \|u\|_2 = 1, \qquad z = ru + w, \quad u^\top w = 0.$$

Then

$$\|z-x\|_2^2 = (r - \|x\|_2)^2 + \|w\|_2^2, \qquad \|z\|_2 \geq |r|.$$

It is suboptimal to have $\|w\|_2 > 0$, so at the optimum $w = 0$ and $z = ru$, $r \geq 0$.

(iii) Substituting $z = ru$ gives

$$\Phi(r) = \tfrac{1}{2}(r - \|x\|_2)^2 + \alpha_j r = \tfrac{1}{2}r^2 - r\|x\|_2 + \alpha_j r + \mathrm{const}.$$

Differentiate w.r.t. r:

$$\frac{d\Phi}{dr} = r - \|x\|_2 + \alpha_j \Longrightarrow r = \|x\|_2 - \alpha_j \quad \text{(stationary)}.$$

Enforcing $r \geq 0$ yields

$$r^* = \max\{0, \|x\|_2 - \alpha_j\}.$$

(iv) Hence

$$z^* = r^*u = \begin{cases} \big(\|x\|_2 - \alpha_j\big)\dfrac{x}{\|x\|_2}, & \|x\|_2 > \alpha_j, \\ 0, & \|x\|_2 \leq \alpha_j. \end{cases}$$

Equivalently,

$$z^* = \max\Big\{0, 1 - \frac{\alpha_j}{\|x\|_2}\Big\}x = \max\Big\{0, 1 - \frac{t\lambda\omega_j}{\|x\|_2}\Big\}x.$$

Putting everything together, we have

$$\big[\mathrm{prox}_{h,t}(x)\big]_{(j)} = \left(1 - \frac{t\lambda\omega_j}{\|x_{(j)}\|_2}\right)^+ x_{(j)}, \quad j = 1, 2, \ldots, J.$$

Algorithms see Jupyter Notebook file.