$x^0 = (x_1^0, \dots, x_n^0)$

ronrd $k+1$

$$x_i^{k+1} = \underset{y \in R}{\arg\min}\, f(x_1^{k+1}, \dots, x_{i-1}^{k+1}, y, x_{i+1}^{k}, \dots, x_n^{k})$$

$\boxed{\text{Dual}}$ method and ADMM.

- Dual methods operate on the dual of a problem that has the form

$$\min_x f(x) \text{ subject to } Ax = b.$$

for convex f. The dual (sub) gradient methods choose an initial $u^{(0)}$ and repeat for $k = 1, 2, 3$.

$$x^{(k)} \in \arg\min f(x) + (u^{(k-1)})^T Ax.$$

$$u^{(k)} = u^{(k-1)} + \underbrace{t_k}_{\text{step size}} (Ax^{(k-1)} - b)$$

Pros : decomposability in the first step.

Con : Poor convergence properties.

→ Can improve convergence by augmenting the Lagrang

Perform blockwise ADMM.

---

minimization tools:

- First-order method.
- Newton's method
- Dual method.
- Interior-point method.

The SDM for unconstrained convex Lipschitz Optimizat

steepest gradient descent.

(first order) Lipschitz condition

a function $f: X \to Y$ is called Lipschitz Continuous if there exists a real constant $k \geq 0$ such that for all x_1, x_2 in X:

$$d_Y(f(x_1), f(x_2)) \leq k\, d_X(x_1, x_2).$$

metric

Lipschitz constant.

adopt the fixed step-size rule: $x^{k+1} = x^k - \frac{1}{\beta} \nabla f(x^k)$

Theorem 1: For convex Lipschitz optimization the Steepest Descent Method generate a sequence of solution such that.

$$f(x^{k+1}) - f(x^*) \le \frac{\beta}{k+2} \|x^0 - x^*\|^2 \text{ and } \min_{i=0,\dots,k} \|\nabla f(x^i)\|^2 = \frac{4\beta^2}{(k+1)(k+2)} \|x^0 - x^*\|^2$$

x^* is the minimize. ✓

ASDM the Accelerated Steepest Descent Method

$$\rightarrow \quad \lambda^0 = 0, \ \lambda^{k+1} = \frac{1+\sqrt{1+4(\lambda^k)^2}}{2}, \ \alpha^k = \frac{1-\lambda^k}{\lambda^{k+1}}.$$

$$\tilde{x}^{k+1} = x^k - \frac{1}{\beta}\nabla f(x^k), \ x^{k+1} = (1-\alpha^k)\tilde{x}^{k+1} + \alpha^k \tilde{x}^k$$

$$\boxed{x^{k+1} = (1-\alpha^k)\left(x^k - \frac{1}{\beta}\nabla f(x^k)\right) + \alpha^k\left(x^{k-1} - \frac{1}{\beta}\nabla f(x^{k-1})\right)}$$

and $(\lambda^k)^2 = \lambda^{k+1}(\lambda^{k+1}-1)$, $\lambda^k > k/2$ and $\alpha^k < 0$

Theorem 2.

$$f(\tilde{x}^{k+1}) - f(x^*) \le \frac{2\beta}{k^2} \|x^0 - x^*\|^2, \ \forall k \ge 1.$$

Convergence: $\delta^{k+1} = f(\tilde{x}^{k+1}) - f(x^*) \le \frac{2\beta}{k^2}\|\Delta^0\|^2$

$$\Delta^k = \lambda^k x^k - (\lambda^k - 1)\tilde{x}^k - x^*$$

First order algorithm: Simply constrain optimization

- Non negative Linear Regression: data $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$

$\min f(x) = \frac{1}{2}\|Ax - b\|^2$ s.t. $x \geq 0$ where $\nabla f(x) = A^T(Ax-b)$

- Semidefinite Linear Regression: Given Data $A_i \in S^n$

for $i = 1, \dots, m$ and $b \in \mathbb{R}^m$

$$\min f(X) = \frac{1}{2}\|Ax - b\|^2 \text{ s.t. } X \succeq 0$$

(p.s. $\succeq$ is typically used in the context of matrices and vectors)

- If used in the context of vectors, it typically means that all elements of the vectors are non negative
- Matrices → non-negative definite)

where $\nabla f(x) = A^T(Ax - b)$

$$AX = \begin{pmatrix} A_1 \cdot X \\ \vdots \\ A_m \cdot X \end{pmatrix} \text{ and } A^T y = \sum_{j=1} y_j A_j$$

---

1 The Logarithmic Barrier Method.

barrier regularization:

$$\min \ \phi(x) = f(x) - \mu B(x)$$

Where $B(x)$ is a barrier function keeping x in the interior of cone k and μ is a small positive parameter.

- $k = R^n_+$ (nonnegative orthant)

$$B(x) = -\sum_j \log(x_j), \ \nabla B(x) = \begin{pmatrix} \frac{-1}{x_1} \\ \cdots \\ \frac{-1}{x_n} \end{pmatrix} \in R^n$$

- $k = S^n_+$ (second-order cone).

$$B(x) = -\log(\det(X)), \ \nabla B(x) = -X^{-1} \in S^n.$$

Newton's method.

$$y = f'(x_n)(x - x_n) + f(x_n)$$

x intercept $y=0$

$$0 = f'(x_n)(x_{n+1} - x_n) + f(x_n)$$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Interior-point methods. pass

Back to Coordinate-wise minimization

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right) = 0$$

$$\rightarrow f(x) = \underbrace{g(x)}_{\substack{\text{convex} \\ \text{diff}}} + \sum_{i=1}^{n} \underbrace{h_i(x_i)}_{\substack{\text{convex} \\ \text{diff}}}$$

Coordinate descent

$$x_1^{(k)} \in \underset{x_1}{\operatorname{argmin}} f(x_1, x_2^{(k-1)} \dots x_n^{(k-1)})$$

$$x_2^{(k)} \in \underset{x_2}{\operatorname{argmin}} f(x_1^{(k-1)}, x_2, \dots, x_n^{(k-1)})$$

$$\vdots$$

→ order is arbitery

One at a time !

Lasso regression

to solve $\min_{\beta_0, \beta} \left\{ \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 \right\}$ s.t $\sum_{j=1}^{p} |\beta_j| \le t$

SVM

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{2} \alpha^T K \alpha - 1^T \alpha \quad \text{s.t. } y^T \alpha = 0, \ 0 \le \alpha \le C1$$

Sequential minimal optimization (SMO)

a basically blockwise coordinate descent

(complementary slackness ⟺ Duality theory)

It is possible to find a solution to the dual problem when only the optimal solution to the primal is known

Largangian dual problem

complementary slackness conditions

$$\alpha_i \cdot [(Av)_i - y_i d - (1 - s_i)] = 0, \quad i = 1, \dots, n \quad (1)$$

$$(C - \alpha_i) \cdot s_i = 0 \quad \quad i = 1, \dots, n \quad (2)$$

v, d, s are primal coefficients, intercept and slacks.

$v = A^T \alpha$, d from (1) with any $0 < \alpha_i < C$ and s computed from (1), (2)

SMO repeats:

1. Choose α_i, α_j that do not satisfy complementary slackness

2. Minimize over α_i, α_j exactly, keeping all other variable fixed.

Standard Convex Optimization

Convex optimization

→ Duality

↓

Lagrangian.

Std: min $f_0(x)$

s.t $f_i(x) \le 0$ $i = 1, \dots m$

$h_i(x) = 0$ $i = 1, \dots, p$

$x \in R$, optimal value p^*, Domain D

→ Lagrangian: $L: R^n \times R^m \times R^p \to R$

with $\operatorname{dom} L = D \times R^m \times R^p$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$