Unconstrained Optimization Algorithms

General Ideas

- Select step size s_k and direction v_k such that $f_0(x_k + s_k v_k)$ is substantially smaller than $f_0(x_k)$.
- v_k is usually selected such that $\nabla f_0(x_k)^\top v_k < 0$
- First order descent

Armijo backtracking step size rule

1. Chosse $\alpha, \beta \in (0,1)$, set $s = 1$
2. If $f_0(x_k + sv_k) \le f_0(x_k) + s\alpha \nabla f_0(x_k)^\top v_k$, then return $s_k = s$.
3. Set $s = \beta s$ and return to step 2.

Newton-type methods

- Newton's method: $v_k = -\nabla^2 f_0(x_k)^{-1} \nabla f_0(x_k)$
- Interpretation: Minimizing second order approximation: $\min_x f_0(x_k) + \nabla f_0(x_k)^\top x + \frac{1}{2} x^\top \nabla^2 f_0(x_k) x$
- Quasi-Newton and variable-metric: $v_k = -H_k \nabla f_0(x_k)$
- Combined with any step size rule

Constrained Optimization Algorithms

Methods

- **Projected Decent:** Use a direction v_k from unconstrained algorithms and compute $x'_{k+1} = x_k + s_k v_k$ and then project x'_{k+1} onto the feasible set.
- Construct an unconstrained problem that approximates the original problem.

Barrier method

- Consider the problem $\min_x f_0(x)$ subject to $f_i(x) \le 0$ for $i = 1...m$.
- Define the approximate problem for some parameter $t \ge 0$: $\min_x f_0(x) + \frac{1}{t}\phi(x)$, where the barrier function $\phi(x)$ tends to ∞ as x tends to the boundary of the feasible set from inside.
- We can use the logarithmic barrier function: $\phi(x) = -\sum_{i=1}^m log(-f_i(x))$
- If f_0, f_i are convex, the approximated problem is convex
- The KKT condition is $t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) = 0$
- The approximation error: $0 \le f_0(x^*(t)) - p^* \le \frac{m}{t}$

Machine Learning

Linear Model

- Linear model: $y(x) = w^\top \phi(x) + e(x)$, where $\phi(x)$ is a feature map and $e(x)$ is a random error
- Least square: $\hat{w} \in \arg\min_w \sum_i (w^\top \phi(x_i) - y_i)^2$
- Quantile for $\alpha \in [0,1)$: $\hat{w} \in \arg\min_w \frac{1}{m(1-\alpha)} \sum_i max\{0, w^\top \phi(x_i) - y_i\} - \frac{1}{m} \sum_i w^\top \phi(x_i) - y_i$. Still a linear program.

Support Vector Machine

- Classification rule: $\hat{y}(x) = sign(\phi(x)^\top x)$
- Convex error function(hinge): $\max\{0, 1-\alpha\}$
- Objective: $\min_w \frac{1}{m} \sum_i \max\{0, 1 - y_i \phi(x_i)^\top w\}$
- We can add L_1 or L_2 norm regularization on w

Logistic Regression

- Variant of SVM with another loss function: $\min_w \frac{1}{m} \sum_i log[1 + exp(-y_i \phi(x_i)^\top w)]$
- Interpretation: $P\{Y = y | X = x\} = \frac{1}{1 + exp(-y\phi(x)^\top w)}$

Fisher discrimination

- Idea: find a direction in $\mathbb{R}^n$, such that the projection of data points onto the direction results in those points with positive label are from those with negative label
- Define $A_+ \in R^{n,m_+}$ is the matrix whose column cosists of x_i such that $y_i = 1$, and A_- similarly for negive labeled points. The centered(mean subtracted) data matrices are $\widetilde{A}_\pm = A_\pm (I_{m_\pm} - \frac{1}{m_\pm} \mathbf{1}\mathbf{1}^\top)$
- The mean-squared variation of the data projected on u around its centroid is: $u^\top M u = u^\top (\frac{1}{m_-} \widetilde{A}_- \widetilde{A}_-^\top + \frac{1}{m_+} \widetilde{A}_+ \widetilde{A}_+^\top) u$
- The goal is: $\max_{u \ne 0} \frac{(u^\top c)^2}{u^\top M u} = \min u^\top M u$ subject to $u^\top c = 1$
- The problem is convex, hence KKT condition implies: $u = \frac{1}{c^\top M^{-1} c} M^{-1} c$

SPCA and NNMF

- The original PCA problem: $\max_z z^\top \widetilde{X} \widetilde{X}^\top z$ s.t. $\|z\|_2 = 1$
- In sparse PCA, we add constraint $card(z) \le k$
- The problem leads to the rank 1 approximation problem: $\min_{p,q} \|X - pq^\top\|_F$ s.t. $card(p) \le k, card(q) \le h$
- One example of rank 1 approximation is: $\min_{p \ge 0, q \ge 0} \|X - pq^\top\|_F$ s.t. $card(p) \le k, card(q) \le h$. A standard solution is coordinate descent.

Finance

Mean-Variance (Markowitz) Models

- $x \in \mathbb{R}^n$ is the wealth allocation in n groups, r is the random vector of unit wealth return for each group. $\hat{r} = E[r], \Sigma = Cov(r)$
- Hence, $\hat{r}^\top x$ is the expected return and $x^\top \Sigma x$ is the variance of the portfolio.
- The problems are usually maximize return with variance constraints and minimize variance, with return constraints.
- Buget constraint: $x^\top \mathbf{1} \le b$, no short-selling: $x \ge 0$
- Sector bounds: $\sum_{i \in S} x_i \le \alpha x^\top \mathbf{1}$ for some $\alpha \in (0,1)$ and sector S.
- Diversification: sum of k largest x_i must not exceed $\eta x^\top \mathbf{1}$ for $\eta \in (0,1)$. This is equivalent to $kt + s^\top \mathbf{1} \le \eta x^\top \mathbf{1}$, $s \ge 0, x - t\mathbf{1} \le s$ where $t \in \mathbb{R}$ and $s \in \mathbb{R}^n$ are additional variables.
- The Sharpe Ratio of a portfolio is $(\hat{r}^\top x - r_f)/\sqrt{x^\top \Sigma x}$, where r_f is the return of a risk-free asset.
- If $\hat{r}^\top x > r_f$ and $\Sigma \succ 0$, then the portfolio that maximizes SR corresponds to a point on the efficient frontier. Max SR can be formulated as SOCP.

Value at Risk