Applications of CLP: Regret Minimization

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye
Regret Minimization: A Convex Optimization Problem

Consider a fractional optimization problem

\[
\min_{(\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^d)} \sum_{i=1}^{n} \frac{(y_i - x_i^T \beta)^2}{z_i^T \alpha}
\]

s.t. \[
\sum_{j=1}^{p} \alpha_j = 1, \quad \alpha_j \geq 0 \quad \forall j = 1, \ldots, p.
\]

Here, given data \( z_i \) is a non-negative and non-zero vector for all \( i \).

For each term \( \frac{(y_i - x_i^T \beta)^2}{z_i^T \alpha} \), one may introduce three scalar variables \( (\gamma_i, \sigma_i, q_i) \).

Then the problem can be reformulated as a conic LP.
Linearly Constrained Optimization Problem

\[ \begin{align*}
\text{minimize} & \quad f(x) \\
\text{(LCOP)} & \quad \text{subject to} \quad Ax = b \\
& \quad x \geq 0.
\end{align*} \]

We assume that \( A \) has full rank and \( f \) is a differentiable convex function.

The KKT conditions:

\[
\begin{align*}
Xs &= 0 \\
Ax &= b \\
-A^T y + \nabla f(x)^T - s &= 0 \\
(x, s) &\geq 0.
\end{align*}
\]
For feasible \((x, y, s) \in \text{int } F\) a primal-dual potential function can be similarly defined by

\[
\psi_{n+\rho}(x, s) := (n + \rho) \log(x^T s) - \sum_{j=1}^{n} \log(x_j s_j).
\]

Once we have an interior feasible point \((x, y, s)\) with \(\mu = x^T s / n > 0\), we can generate a new iterate \((x^+, y^+, s^+)\) by solving for \((d_x, d_y, d_s)\) from the system of linear equations:

\[
\begin{align*}
S d_x + X d_s &= \frac{x^T s}{n+\rho} \cdot e - X s, \\
A d_x &= 0, \\
-A^T d_y + \nabla^2 f(x) d_x - d_s &= 0.
\end{align*}
\]

Note that \(d_x^T d_s \geq 0\) here, and one can choose \(\rho = O(\sqrt{n})\).
Let the direction \( \mathbf{d} = (d_x, d_y, d_s) \) be generated by equation (3), and let

\[
\theta = \frac{\alpha \sqrt{\min(Xs)}}{\|(XS)^{-1/2}(\frac{x^T s}{n+\rho}e - Xs)\|},
\]

where \( \alpha \) is a positive constant less than 1. Let

\[
x^+ = x + \theta d_x, \quad y^+ = y + \theta d_y, \quad \text{and} \quad s^+ = s + \theta d_s.
\]

Then, we have \((x^+, y^+, s^+) \in \text{int } \mathcal{F}\) and

\[
\psi_{n+\rho}(x^+, s^+) - \psi_{n+\rho}(x, s) \leq -\delta,
\]

for a positive constant \( \delta \) with a suitable choice of \( \alpha \).
Theorem 1  Let \( \rho = O(\sqrt{n}) \) and \( f(x) \) satisfy a technical condition. Then, the algorithm terminates in at most \( O(\sqrt{n} \log((x^0)^T s^0 / \epsilon)) \) iterations with

\[
(x^k)^T s^k \leq \epsilon.
\]
Once we have an interior feasible point \((x, y, s)\) with \(\mu = x^T s/n > 0\), we can generate a new iterate \((x^+, y^+, s^+)\) by solving for \((d_x, d_y, d_s)\) from the system of linear equations:

\[
\begin{align*}
Sd_x +Xd_s &= \gamma \mu \cdot e - Xs, \\
Ad_x &= 0, \\
-A^T d_y + \nabla^2 f(x) d_x - d_s &= 0;
\end{align*}
\]

where \(\gamma\) is a constant between 0 and 1.

Then we have

\[
\begin{align*}
x^+ &= x + \alpha d_x, \\
y^+ &= y + \alpha d_y, \\
s^+ &= \nabla f(x^+) - A^T y^+,
\end{align*}
\]

where stepsize \(\alpha\) is set to ensure interior.
Once we have an interior feasible point \((x, y, s)\) with \(\mu = x^T s / n > 0\), we can generate a new iterate \((x^+, y^+, s^+)\) by solving for \((d_x, d_y, d_s)\) from the system of linear equations:

\[
\begin{align*}
\mu \cdot X^{-2} d_x + d_s &= \gamma \mu \cdot e - X s, \\
A d_x &= 0, \\
-A^T d_y + \nabla^2 f(x) d_x - d_s &= 0;
\end{align*}
\]

where again \(\gamma\) is a constant between 0 and 1.

Then we have

\[
x^+ = x + \alpha d_x, \quad y^+ = y + \alpha d_y, \quad s^+ = \nabla f(x^+) - A^T y^+,
\]

where stepsize \(\alpha\) is set to ensure interior.
A Conic LP Reformulation

\[
\begin{align*}
\min_{(\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^d, \gamma, \sigma, q)} & \quad \sum_{i=1}^{n} \gamma_i \\
\text{s.t.} & \quad \sum_{j=1}^{p} \alpha_j = 1, \\
& \quad z_i^T \alpha - q_i = 0, \quad \forall i, \\
& \quad x_i^T \beta - \sigma_i = y_i, \quad \forall i, \\
& \quad \alpha_j \geq 0, \quad \forall j, \\
& \quad \begin{pmatrix} \gamma_i & \sigma_i \\ \sigma_i & q_i \end{pmatrix} \succeq 0, \quad \forall i.
\end{align*}
\]

This is a standard (dual) conic linear optimization problem with a product of many \(2 \times 2\) positive semi-definite matrix cones, and a non-negative cones on \(\alpha\).
Validation of the Reformulation

The positive semi-definiteness of the $2 \times 2$ matrix means $\gamma_i q_i \geq \sigma_i^2$ so that

$$\gamma_i \geq \frac{\sigma_i^2}{q_i} = \frac{(y_i - x_i^T \beta)^2}{z_i^T \alpha}.$$ 

Since $\gamma_i$ is a upper bound variable on each term, minimizing $\sum_i \gamma_i$ equivalent to minimizing

$$\sum_i \frac{(y_i - x_i^T \beta)^2}{z_i^T \alpha}.$$ 

The new formulation has $2n + 1$ equality constraints with two-block structures (one involving $(\alpha, q)$ and the other $(\beta, \sigma)$). One may add more linear constraints to link $\alpha$ and $\beta$, and it becomes a standard conic linear optimization problem.
Standard Dual CLP Form

\[
\max_{(\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^d, \gamma \in \mathbb{R}^n)} \quad -\sum_{i=1}^{n} \gamma_i \\
\text{s.t.} \\
\begin{pmatrix}
-\gamma_i & x_i^T \beta \\
x_i^T \beta & -z_i^T \alpha
\end{pmatrix} \preceq \begin{pmatrix}
0 & y_i \\
y_i & 0
\end{pmatrix}, \forall i,
\]
\[
\sum_{j=1}^{p} \alpha_j \leq 1, \\
-\alpha_j \leq 0, \forall j.
\]

Note that the left-hand-side of the matrix inequality can be written as

\[
\sum_{j=1}^{p} \alpha_j \begin{pmatrix}
0 & 0 \\
0 & -z_{ij}
\end{pmatrix} + \sum_{j=1}^{d} \beta_j \begin{pmatrix}
0 & x_{ij} \\
x_{ij} & 0
\end{pmatrix} + \gamma_i \begin{pmatrix}
-1 & 0 \\
0 & 0
\end{pmatrix}.
\]

It would be solved by a standard conic linear programming solvers, such as SEDUMI and DSDP.
Augmented Lagrangian for the Dual CLP with Mixed Cones

$$\min_{(y \in \mathbb{R}^m, s)} \quad -b^T y$$

s.t. \quad \begin{align*}
A_p^T y + s_p &= c_p, \quad p = 1, \ldots, P, \\
s_p &\in C_p,
\end{align*}

where each $C_p$ is a closed convex cone.

$$L(y, s, x) = -b^T y - \sum_{p=1}^{P} (x_p \cdot (A_p^T y + s_p - c_p)) + \frac{\beta}{2} \left( \sum_{p=1}^{P} ||A_p^T y + s_p - c_p||^2 \right),$$

where now $x_p$ is the multiplier of the $p$th equality constraint.
**ADMM for the Dual CLP with Mixed Cones**

Then, for any given \((y^k, s^k, x^k)\), we compute a new iterate pair

\[
y^{k+1} = \arg \min_y L(y, s^k, x^k)
\]

\[
s^{k+1} = \arg \min_{s \in C} L(y^{k+1}, s, x^k)
\]

and

\[
x_{p}^{k+1} = x_{p}^k - \beta (A_{p}^T y^{k+1} + s_{p}^{k+1} - c_{p}), \quad p = 1, \ldots, P.
\]

The minimization of cone variables \(s\) can be implemented in parallel:

\[
s_{p}^{k+1} = \arg \min_{s_{p} \in C_{p}} \left( x_{p}^k \cdot (A_{p}^T y^{k+1} + s_{p} - c_{p}) + \frac{\beta}{2} \| A_{p}^T y^{k+1} + s_{p} - c_{p} \|^2 \right),
\]

since they are minimized independently.
Affine-Scaling Augmented Lagrangian

For any given \((y^k, s^k \in \text{int } C, x^k)\), we consider the Affine-Scaling Augmented Lagrangian Function

\[
L^k(y, s, x) = \]

\[
- b^T y - \sum_{p=1}^{P} (x_p \cdot (S_p^k)^{-1}(A_p^T y + s_p - c_p)) + \frac{\beta}{2} \left( \sum_{p=1}^{P} \|(S_p^k)^{-1}(A_p^T y + s_p - c_p)\|^2 \right).
\]
Then the affine-scaling ADMM would be

\[
y^{k+1} = \arg \min_y L^k(y, s^k, x^k)
\]

\[
s^* = \arg \min_s L^k(y^{k+1}, s, x^k)
\]

and update

\[
s^{k+1} = \alpha s^k + (1 - \alpha)s^*_2
\]

where step-size \( \alpha \) is chosen such that \( s^{k+1} \) remains in the interior of \( C \);

\[
x^{k+1}_p = x^k_p - \beta(A_p^T y^{k+1} + s^{k+1}_p - c_p), \quad p = 1, \ldots, P.
\]