

Quadratic Programming

0 Reference

1. Thomas Reslow Krüth, Interior-Point Algorithms for Quadratic Programming, Kongens Lyngby 2008, IMM-M.Sc-2008-19

以下この論文を参考に、制約条件の不等号を逆にして書く．

1 主問題

主問題は以下のようになる．

$$\begin{aligned} \text{(P)} : \quad & \text{Minimize } \frac{1}{2} \mathbf{x}^\top G \mathbf{x} + \mathbf{g}^\top \mathbf{x} \\ & \text{subject to } A^\top \mathbf{x} \preceq \mathbf{b} \end{aligned} \quad (1)$$

但し、最適化変数を $\mathbf{x} = [x_1 \cdots x_n]^\top$ としている．係数行列 Q は半正定値行列を仮定している．また、制約の個数を m としている．

2 ラグランジアン・最適性条件

2.1 ラグランジアン

上記主問題より、ラグランジュ乗数 $\boldsymbol{\lambda}$ を用いて、ラグランジアン $L(\mathbf{x}, \boldsymbol{\lambda})$ は以下のように定義できる．

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \mathbf{x}^\top G \mathbf{x} + \mathbf{g}^\top \mathbf{x} + \sum_{i=1}^m \lambda_i (\mathbf{a}_i^\top \mathbf{x} - b_i) \quad (2)$$

制約を破る場合 ($\mathbf{a}_i^\top \mathbf{x} - b_i \geq 0$) にラグランジアンに対してペナルティとして働くようにプラスマイナスを考えればよい ($\lambda_i > 0$ としている)．

2.2 1次最適性条件

これより、1次最適性条件 (KKT) は以下のように書ける．

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) &= G \mathbf{x} + \mathbf{g} + A \boldsymbol{\lambda} = \mathbf{0} \\ A^\top \mathbf{x} - \mathbf{b} &\leq \mathbf{0} \\ \lambda_i (\mathbf{b} - A^\top \mathbf{x})_i &= 0 \\ \boldsymbol{\lambda} &\geq \mathbf{0} \end{aligned} \quad (3)$$

スラック変数 \mathbf{s} を導入して不等式制約を等式制約に直して、

$$\begin{aligned} G \mathbf{x} + \mathbf{g} + A \boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{s} + A^\top \mathbf{x} - \mathbf{b} &= \mathbf{0} \\ s_i \lambda_i &= 0 \\ (\boldsymbol{\lambda}, \mathbf{s}) &\geq \mathbf{0} \end{aligned} \quad (4)$$

このシステムを以下のように書き直す.

$$F(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = \begin{bmatrix} G\mathbf{x} + \mathbf{g} + A\boldsymbol{\lambda} \\ \mathbf{s} + A^\top \mathbf{x} - \mathbf{b} \\ S\boldsymbol{\Lambda} \mathbf{e} \end{bmatrix} = \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} \end{bmatrix} \quad (5)$$

2.3 Newton 方向

1 次最適性条件から得られたシステム

$$F(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = \mathbf{0} \quad (6)$$

の解を求める. そこで Newton 法を利用して, 二次近似の最小値方向に下っていく.

Newton 方向は, F のヤコビアンを J として,

$$J(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) = \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} = -F(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s}) \quad (7)$$

の解として与えられる. これを書き下せば以下を得る.

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \boldsymbol{\Lambda} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} \end{bmatrix} \quad (8)$$

あとはこの Newton 方向と直線探索を利用して, 適切なステップ幅 α を用いて,

$$\begin{bmatrix} \mathbf{x}_{k+1} \\ \boldsymbol{\lambda}_{k+1} \\ \mathbf{s}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\lambda}_k \\ \mathbf{s}_k \end{bmatrix} + \alpha \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} \quad (9)$$

のように更新する.

3 システムの解き方

実際にシステム (8) を解く際には逆行列は計算せず, コレスキー分解を利用する. これは後で述べる予測ステップ・修正ステップにおいて, システム (8) をそれぞれ解かなければならないが, 左辺は変わらないため, コレスキー分解を利用して解いたほうが計算量が抑えられる.

以下のシステムを解くことを考える.

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \boldsymbol{\Lambda} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} \end{bmatrix} \quad (10)$$

2 段目から,

$$\Delta \mathbf{s} = -r_p - A^\top \Delta \mathbf{x} \quad (11)$$

3 段目から, 上式を利用して,

$$\begin{aligned} \Delta \boldsymbol{\lambda} &= -S^{-1}(\boldsymbol{\Lambda} \Delta \mathbf{s} + r_{s\lambda}) \\ &= S^{-1}(-r_{s\lambda} + \boldsymbol{\Lambda} r_p) + S^{-1} \boldsymbol{\Lambda} A^\top \Delta \mathbf{x} \end{aligned} \quad (12)$$

更に 1 段目から,

$$\begin{aligned} -r_d &= G\Delta \mathbf{x} + A\Delta \boldsymbol{\lambda} \\ &= (G + AS^{-1}\boldsymbol{\Lambda}A^\top)\Delta \mathbf{x} + AS^{-1}(-r_{s\lambda} + \boldsymbol{\Lambda}r_p) \\ &= \bar{G}\Delta \mathbf{x} + \bar{r} \end{aligned} \quad (13)$$

但し,

$$\bar{G} = G + AS^{-1}\boldsymbol{\Lambda}A^\top \quad (14)$$

$$\bar{r} = AS^{-1}(-r_{s\lambda} + \boldsymbol{\Lambda}r_p) \quad (15)$$

従って、線形方程式

$$\bar{G}\Delta\mathbf{x} = -(r_d + \bar{r}) \quad (16)$$

を解けば、

$$\Delta\mathbf{s} = -r_p - A^\top \Delta\mathbf{x} \quad (17)$$

$$\Delta\boldsymbol{\lambda} = -S^{-1}(\Lambda\Delta\mathbf{s} + r_{s\lambda}) \quad (18)$$

として $\mathbf{s}, \boldsymbol{\lambda}$ も求められる。

線形方程式 (16) を解く際にコレスキー分解を利用し、下三角行列 L を利用して、 \bar{G} は以下のように分解できる。

$$\bar{G} = LL^\top \quad (19)$$

そして以下の方程式を解けばよい。

$$LY = -(r_d + \bar{r}) \quad (20)$$

$$L^\top \Delta\mathbf{x} = Y \quad (21)$$

プログラム上では一反復においてこの作業を 2 回することになるが、 \bar{G} のコレスキー分解を保存できるので、計算量を減らすことができる。

4 Predictor-Corrector Method

基本的なアイデアはシステム (8) に適切な修正を加えながら解くイメージ。

4.1 中心曲線

中心曲線はスカラーパラメータ τ によって形成される、狭義の実行可能点であり、以下で表される。

$$F(\mathbf{x}_\tau, \boldsymbol{\lambda}_\tau, \mathbf{s}_\tau) = \begin{bmatrix} 0 \\ 0 \\ \tau \mathbf{e} \end{bmatrix}, \quad (\boldsymbol{\lambda}_\tau, \mathbf{s}_\tau) > \mathbf{0} \quad (22)$$

中心曲線上の点は狭義に実行可能であり、 $r_d = 0, r_p = 0$ を満たす。

各反復ごとに $(\mathbf{x}_k, \boldsymbol{\lambda}_k, \mathbf{s}_k)$ が中心曲線に沿って進み、 τ が各ステップで減少するようにすればよい。

更に以下の双対性測度 μ 、中心化パラメータ $\sigma \in [0, 1]$ を定義する。

$$\mu = \frac{\mathbf{s}^\top \boldsymbol{\lambda}}{m} \quad (23)$$

μ は $s_i \lambda_i$ についての情報を持ち、計算されたステップが $\mathbf{s}^\top \boldsymbol{\lambda}$ を大幅に減少させたかどうかを判断することで、計算された探索方向がどれだけ有用であるかを示す。例えば、 $\mathbf{s}^\top \boldsymbol{\lambda}$ が大幅に減少したなら、中心化はほぼ必要なく ($\sigma = 0$)、あまり減少しないなら、計算された探索方向はあまり有用ではなく、大きな中心化 ($\sigma = 1$) が必要となる。

4.2 予測ステップ

実際に今の反復において、現在地がどの程度中心曲線から離れているかを知るために、以下に示す予測ステップを行う。

予測ステップではアフィンスケーリング方向 $(\Delta\mathbf{x}^{\text{aff}}, \Delta\boldsymbol{\lambda}^{\text{aff}}, \Delta\mathbf{s}^{\text{aff}})$ を得るために以下を計算し、さらに後に示す直線探索によって α^{aff} を計算する。

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}^{\text{aff}} \\ \Delta\boldsymbol{\lambda}^{\text{aff}} \\ \Delta\mathbf{s}^{\text{aff}} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} \end{bmatrix} \quad (24)$$

求めたアフィンスケーリング方向 $(\Delta \mathbf{x}^{\text{aff}}, \Delta \boldsymbol{\lambda}^{\text{aff}}, \Delta \mathbf{s}^{\text{aff}})$ とステップ幅 α^{aff} によって、予測双対性測度 μ^{aff} が計算できる。

$$\mu^{\text{aff}} = \frac{(\mathbf{s} + \alpha^{\text{aff}} \Delta \mathbf{s}^{\text{aff}})^\top (\boldsymbol{\lambda} + \alpha^{\text{aff}} \Delta \boldsymbol{\lambda}^{\text{aff}})}{m} \quad (25)$$

μ^{aff} と現在の双対性測度 μ を比較することで、アフィンスケーリング方向が良い方向なのか判断できる。

実際は中心化パラメータ σ は以下のように決められる。

$$\sigma = \left(\frac{\mu^{\text{aff}}}{\mu} \right)^3 \quad (26)$$

中心化パラメータ σ は予測ステップで得られた探索方向をどの程度中心に寄せる必要があるかを反映するように選ばれる。極端な例として、 $\sigma = 0$ なら、中心化する必要はなく、標準的な Newton ステップが採用される。 $\sigma = 1$ なら、ステップは中心に寄せられ（すなわち、反復点を中心曲線に近づけるために使用され）、双対性測度を減少させる進展はほぼないが、次の反復でより長いステップを採れる（ $\lambda_i s_i$ の値を減らして、相補性条件を満たす場所に近づけるイメージ?）。

4.3 ステップ幅計算

$(\boldsymbol{\lambda}, \mathbf{s}) > \mathbf{0}$ を破らないように α を決める。

$$\boldsymbol{\lambda} + \alpha \Delta \boldsymbol{\lambda} \geq \mathbf{0} \quad (27)$$

$$\mathbf{s} + \alpha \Delta \mathbf{s} \geq \mathbf{0} \quad (28)$$

$$(29)$$

これより、例えば $\boldsymbol{\lambda}$ に対しては以下のように選択する。

- $\Delta \lambda_i > 0$: $\alpha_\lambda = 1$
- $\Delta \lambda_i = 0$: $\alpha_\lambda = 1$
- $\Delta \lambda_i < 0$: $\lambda_i + \alpha_\lambda \Delta \lambda_i = 0 \Rightarrow \alpha_\lambda = -\frac{\lambda_i}{\Delta \lambda_i}$

これより α は以下のように選べばよい。

$$\alpha_\lambda = \min_{i: \Delta \lambda_i < 0} \left(1, \min -\frac{-\lambda_i}{\Delta \lambda_i} \right) \quad (30)$$

$$\alpha_s = \min_{i: \Delta s_i < 0} \left(1, \min -\frac{-s_i}{\Delta s_i} \right) \quad (31)$$

$$\alpha = \min(\alpha_\lambda, \alpha_s) \quad (32)$$

4.4 修正ステップ

まずは以下のシステムを解くところから始める。

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^{\text{aff}} \\ \Delta \boldsymbol{\lambda}^{\text{aff}} \\ \Delta \mathbf{s}^{\text{aff}} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} \end{bmatrix} \quad (33)$$

このシステムによって求められるアフィンスケーリング方向 $(\Delta \mathbf{x}^{\text{aff}}, \Delta \boldsymbol{\lambda}^{\text{aff}}, \Delta \mathbf{s}^{\text{aff}})$ に進むと、

$$(s_i + \Delta s_i^{\text{aff}})(\lambda_i + \Delta \lambda_i^{\text{aff}}) = s_i \lambda_i + s_i \Delta \lambda_i^{\text{aff}} + \lambda_i \Delta s_i^{\text{aff}} + \Delta s_i^{\text{aff}} \Delta \lambda_i^{\text{aff}} \quad (34)$$

$$= \Delta s_i^{\text{aff}} \Delta \lambda_i^{\text{aff}} \quad (35)$$

となり、 $\Delta s_i^{\text{aff}} \Delta \lambda_i^{\text{aff}} \neq 0$ なので、アフィンスケーリングを実行すると、線形化エラーとなる（変形では三段目の $S \Delta \lambda_i^{\text{aff}} + \Lambda \Delta s_i^{\text{aff}} = -S \Lambda \mathbf{e}$ を用いている）。

アフィンスケーリングを実行しても相補性が保たれるように、右辺に以下のように $\Delta s_i^{\text{aff}} \Delta \lambda_i^{\text{aff}}$ を加えておけばよい。

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^{\text{cor}} \\ \Delta \boldsymbol{\lambda}^{\text{cor}} \\ \Delta \mathbf{s}^{\text{cor}} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} + \Delta S^{\text{aff}} \Delta \Lambda^{\text{aff}} \mathbf{e} \end{bmatrix} \quad (36)$$

実際に解くのは、これにさらに中心化を加えた、以下のシステムであり、ここから求めた Newton 方向が実際に採用する探索方向である。

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} + \Delta S^{\text{aff}} \Delta \Lambda^{\text{aff}} \mathbf{e} - \sigma \mu \mathbf{e} \end{bmatrix} \quad (37)$$

更に、この Newton 方向に対して、適切なステップ幅 α を前述の方法で求め状態を更新するが、その際

$$\eta = 0.95 \quad (38)$$

程度の減速パラメータ η を用いて、

$$\begin{bmatrix} \mathbf{x}_{k+1} \\ \boldsymbol{\lambda}_{k+1} \\ \mathbf{s}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\lambda}_k \\ \mathbf{s}_k \end{bmatrix} + \eta \alpha \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} \quad (39)$$

とすると、良い収束が得られることが確認されているらしい。

5 終了条件

以下の実行可能性・最適性を表す量

$$r_d = G\mathbf{x} + \mathbf{g} + A\boldsymbol{\lambda} \quad (40)$$

$$r_p = \mathbf{s} + A^\top \mathbf{x} - \mathbf{b} \quad (41)$$

$$\mu = \frac{\mathbf{s}^\top \boldsymbol{\lambda}}{m} \quad (42)$$

に対して、十分小さな正の定数 ε を用いて

$$\|r_d\| \leq \varepsilon \quad (43)$$

$$\|r_p\| \leq \varepsilon \quad (44)$$

$$|\mu| \leq \varepsilon \quad (45)$$

$$(46)$$

が満たされるとき、終了とする。

6 アルゴリズム

Step 1: 初期値 $(\mathbf{x}_0, \boldsymbol{\lambda}_0, \mathbf{s}_0)$ を決定。(実行可能解である必要はないが、 $(\boldsymbol{\lambda}, \mathbf{s}) \geq \mathbf{0}$ となる点を選ぶ)

Step 2: 残差、双対性測度を計算する。

$$r_d = G\mathbf{x}_0 + \mathbf{g} + A\boldsymbol{\lambda}_0$$

$$r_p = \mathbf{s}_0 + A^\top \mathbf{x}_0 - \mathbf{b}$$

$$r_{s\lambda} = S_0 \Lambda_0 \mathbf{e}$$

$$\mu = \frac{\mathbf{s}_0^\top \boldsymbol{\lambda}_0}{m}$$

Step 3: 予測ステップ

Step 3-1: アフィンスケーリング方向 $(\Delta \mathbf{x}^{\text{aff}}, \Delta \boldsymbol{\lambda}^{\text{aff}}, \Delta \mathbf{s}^{\text{aff}})$, 中心化パラメータ σ を得るために以下を解く。

$$\begin{bmatrix} G & A & O \\ A^\top & O & I \\ O & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}^{\text{aff}} \\ \Delta \boldsymbol{\lambda}^{\text{aff}} \\ \Delta \mathbf{s}^{\text{aff}} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} \end{bmatrix}$$

Step 3-2: ステップ幅を α^{aff} 計算.

$$\begin{aligned}\alpha_{\lambda}^{\text{aff}} &= \min_{i: \Delta \lambda_i^{\text{aff}} < 0} \left(1, \min -\frac{-\lambda_i}{\Delta \lambda_i^{\text{aff}}} \right) \\ \alpha_s^{\text{aff}} &= \min_{i: \Delta s_i^{\text{aff}} < 0} \left(1, \min -\frac{-s_i}{\Delta s_i^{\text{aff}}} \right) \\ \alpha^{\text{aff}} &= \min(\alpha_{\lambda}^{\text{aff}}, \alpha_s^{\text{aff}})\end{aligned}$$

Step 3-3: 予測双対性測度 μ^{aff} を計算.

$$\mu^{\text{aff}} = \frac{(\mathbf{s} + \alpha^{\text{aff}} \Delta \mathbf{s}^{\text{aff}})^{\top} (\boldsymbol{\lambda} + \alpha^{\text{aff}} \Delta \boldsymbol{\lambda}^{\text{aff}})}{m}$$

Step 3-4: 中心化パラメータ σ を計算.

$$\sigma = \left(\frac{\mu^{\text{aff}}}{\mu} \right)^3$$

Step 4: 中心化・修正ステップ

Step 4-1: 探索方向 $(\Delta \mathbf{x}, \Delta \boldsymbol{\lambda}, \Delta \mathbf{s})$ を以下の式を解くことによって得る.

$$\begin{bmatrix} G & A & O \\ A^{\top} & O & I \\ O & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix} = - \begin{bmatrix} r_d \\ r_p \\ r_{s\lambda} + \Delta S^{\text{aff}} \Delta \Lambda^{\text{aff}} \mathbf{e} - \sigma \mu \mathbf{e} \end{bmatrix}$$

Step 4-2: ステップ幅を α^{aff} 計算.

$$\begin{aligned}\alpha_{\lambda} &= \min_{i: \Delta \lambda_i < 0} \left(1, \min -\frac{-\lambda_i}{\Delta \lambda_i} \right) \\ \alpha_s &= \min_{i: \Delta s_i < 0} \left(1, \min -\frac{-s_i}{\Delta s_i} \right) \\ \alpha &= \min(\alpha_{\lambda}, \alpha_s)\end{aligned}$$

Step 5: $(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{s})$ を更新.

$$\begin{bmatrix} \mathbf{x}_{k+1} \\ \boldsymbol{\lambda}_{k+1} \\ \mathbf{s}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\lambda}_k \\ \mathbf{s}_k \end{bmatrix} + \eta \alpha \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \\ \Delta \mathbf{s} \end{bmatrix}$$

Step 6: 残差, 双対性測度を更新.

$$\begin{aligned}r_d &= G\mathbf{x} + \mathbf{g} + A\boldsymbol{\lambda} \\ r_p &= \mathbf{s} + A^{\top} \mathbf{x} - \mathbf{b} \\ r_{s\lambda} &= S\Lambda \mathbf{e} \\ \mu &= \frac{\mathbf{s}^{\top} \boldsymbol{\lambda}}{m}\end{aligned}$$

Step 7: 終了条件が満たされるか判断し, 満たされれば \mathbf{x} を最適解として出力. 満たされなければ **Step 3** に戻る.

$$\begin{aligned}\|r_d\| &\leq \varepsilon \\ \|r_p\| &\leq \varepsilon \\ |\mu| &\leq \varepsilon\end{aligned}$$