Non-Convex Quadratic Optimization with Random Projection (ランダム射影を用いた非凸二次最適化法)

数理情報学専攻 48206229　**富士 晃成**
指導教員　**武田 朗子 教授**

1　はじめに

ランダム射影は高次元の点をそのノルムを大きく変化させることなく低次元へ射影することが知られており [2], 数値データの圧縮に多く用いられるが，近年は最適化手法にも用いられている．D'Ambrosio ら [1] は二次計画問題にランダム射影を応用し，変数をランダムに集約することでサイズの小さな問題に近似的に帰着できることを示し，また Gower ら [3] は探索方向をランダムな部分空間に限定することで各反復の計算量を小さくしたニュートン型の反復法を提案している．しかし，両者ともに実用性の面では凸な問題を対象としており，とくに後者に関しては理論保証は目的関数が凸な場合にしか与えられていない．本研究ではそれぞれの手法を非凸な場合に拡張し，二次計画問題に関しては得られる最適解の誤差の保証，ニュートン型反復法に関しては大域的な収束性という理論的な保証を与えた．それぞれの内容について，既存研究と本研究の特徴をまとめたものを表 1 に示す．

表 1: 既存研究および本研究

(a) 二次計画問題

既存研究 [1]	本研究
小さな問題に帰着 誤差保証あり	**凸**かつ小さな問題に帰着 誤差保証あり

(b) ニュートン型反復法

既存研究 [3]	本研究
目的関数が凸 部分空間を利用 大域的線形収束	目的関数が**非凸** 部分空間 + 正則化を利用 大域的収束

2　ランダム射影

$n, s(s < n)$ をそれぞれ正整数として各成分 P_{ij} が正規分布 $\mathrm{N}(0, 1/s)$ に独立に従う行列 $P \in \mathbb{R}^{s\times n}$ をランダム行列と呼び，対応する線形写像をランダム射影と呼ぶ．このとき，ある定数 $\mathcal{C}_0$ が存在し任意の $x \in \mathbb{R}^n$ と $\varepsilon > 0$ に対して，少なくとも確率 $1 - 2\exp(-\mathcal{C}_0\varepsilon^2 s)$ で

$$(1-\varepsilon)\left\|x\right\|_2^2 \leq \left\|Px\right\|_2^2 \leq (1+\varepsilon)\left\|x\right\|_2^2$$

が成立することが知られている [2]．つまりランダム射影は高確率でベクトルの大きさをほぼ保存する写像であり，これを利用して高次元の点をそのノルムを大きく変化させることなく低次元へ埋め込むことができる．

3　非凸二次計画問題へのランダム射影の応用

3.1　既存研究

$Q \in \mathbb{R}^{n\times n}, c \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ として次の二次計画問題を考える:

$$\mathbf{P} \equiv \min_{x\in\mathbb{R}^n} \{x^\top Qx + c^\top x \mid Ax \leq b\}.$$

D'Ambrosio ら [1] は，ランダム行列 $P \in \mathbb{R}^{s\times n}$ を用いて

$$\mathbf{RP} \equiv \min_{u\in\mathbb{R}^s} \{u^\top PQP^\top u + (Pc)^\top u \mid AP^\top u \leq b\}$$

と書けるランダム射影された問題を代わりに解くことを提案し，各計画問題の最適値を $\mathrm{opt}(\cdot)$ で表すことにすれば $\mathrm{opt}(\mathbf{P}) \simeq \mathrm{opt}(\mathbf{RP})$ となることを示した．ただし $\mathbf{P}$ が非凸な問題であれば $\mathbf{RP}$ も一般には非凸な問題になるので，サイズが小さくなったとしても $\mathbf{RP}$ が容易に解けるわけではない．

3.2　本研究

元問題 $\mathbf{P}$ が非凸な問題であれば $\mathbf{RP}$ も非凸二次計画問題となるが，まず $\mathbf{RP}$ の凸性を決定する二次の係数行列 $PQP^\top$ の固有値分布について考察する．簡単な計算から $\mathrm{E}[PQP^\top] = (\mathrm{tr}\, Q/s)I_s$ がわかるので $PQP^\top$ の固有値は $\mathrm{tr}\, Q/s$ の近くに分布することが期待される．つまり $\mathrm{tr}\, Q > 0$ であれば，図 1 に一例が示されているように，$PQP^\top$ の固有値の多くは正の値をとり $PQP^\top$ が近似的に半正定値行列となる．そこで半正定値錐への射影 $\mathcal{F}^+$ を用いて $\mathbf{RP}$ の目的関数を凸化した次の問題

$$\mathbf{CRP} \equiv \min_{u\in\mathbb{R}^s} \{u^\top \mathcal{F}^+(PQP^\top)u + (Pc)^\top u \mid AP^\top u \leq b\}$$

を考えれば，$\mathbf{CRP}$ は凸二次計画問題で，先行研究の結果と合わせて $\mathrm{opt}(\mathbf{CRP}) \simeq \mathrm{opt}(\mathbf{RP}) \simeq \mathrm{opt}(\mathbf{P})$ が期

待される．本研究で実際に示した誤差保証に関する定理を下に述べる．

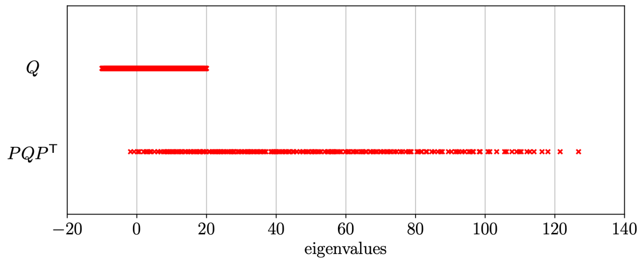

図 1: $Q \in \mathbb{R}^{2000\times 2000}$ と $PQP^\top \in \mathbb{R}^{200\times 200}$ の固有値分布の例

定理. 問題 **CRP** と問題 **P** の最適値の間に常に

$$\mathrm{opt}(\mathbf{CRP}) \geq \mathrm{opt}(\mathbf{P})$$

が成り立つ．また任意の $\varepsilon, \varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ と $\delta_1, \delta_2 \in (0, 1/2)$ に対して，s を $Q, n, m, \varepsilon, \varepsilon_1, \varepsilon_2, \varepsilon_3, \delta_1, \delta_2$ に依存するある条件をみたすように適切にとれば，少なくとも確率 $1 - \delta_1 - \delta_2$ で次の不等式が成り立つ．

$$\alpha\left(1 + \frac{3\varepsilon_1 + 2\varepsilon_2 + \varepsilon_3}{\cos\theta^*}\right)\mathrm{opt}(\mathbf{P}) \geq \mathrm{opt}(\mathbf{CRP}).$$

ここで x^* を問題 **P** の最適解の 1 つとして α は次のように定義される正数である．

$$\alpha = 1 - \frac{\varepsilon\|x^*\|_2}{r + \varepsilon\|x^*\|_2}.$$

また θ^* は $(x^* {x^*}^\top, x^*), (Q, c)$ をそれぞれ $\mathbb{R}^{n^2+n}$ の元とみなした時にそれらがなす角である．

4 反復法へのランダム射影の応用

4.1 既存研究

$f : \mathbb{R}^n \to \mathbb{R}$ を二回連続微分な関数として次の無制約最適化問題を考える:

$$\min_{x\in\mathbb{R}^n} f(x).$$

Gower ら [3] は f が凸な場合にこの問題に対する反復法として次の更新式で与えられるランダム部分空間ニュートン法を提案している．

$$d_k^{\mathrm{RSN}} = -P_k^\top (P_k H_k P_k^\top)^{-1} P_k g_k,$$
$$x_{k+1} = x_k + \frac{1}{\hat{L}} d_k^{\mathrm{RSN}}.$$

ここで，$H_k = \nabla^2 f(x_k), g_k = \nabla f(x_k)$ で $\hat{L}$ は定数，$P_k \in \mathbb{R}^{s\times n}$ は各反復でサンプリングされるランダム行列である．古典的なニュートン法では H_k の逆行列の計算が必要になるため n が大きい時には計算量が非常に大きくなってしまうが，ランダム部分空間ニュートン法では $P_k H_k P_k^\top \in \mathbb{R}^{s\times s}$ の逆行列を計算すれば十分なため一反復あたりの計算量を抑えられる．またランダム部分空間ニュートン法が適切な仮定の元で最適解へ大域的に線形収束することも示されている．しかし f が非凸な場合はそもそも探索方向 d_k^{RSN} が降下方向である保証はない．

4.2 本研究

非凸な目的関数に対して提案されている正則化ニュートン法 [4] を元に，本研究では $c_1 > 1, c_2 > 0, \gamma > 0$ を定数として次の更新式で与えられるランダム部分空間正則化ニュートン法を提案する．

$$\Lambda_k = \max(0, -\lambda_{\min}(P_k H_k P_k^\top))$$
$$d_k = -P_k^\top (P_k H_k P_k^\top + c_1 \Lambda_k I_s + c_2 \|g_k\|_2^\gamma I_s)^{-1} P_k g_k,$$
$$x_{k+1} = x_k + t_k d_k.$$

ステップサイズ t_k はアルミホルールおよびバックトラッキングによって決定する．ランダム部分空間正則化ニュートン法の理論保証として，本研究では次のような大域的収束性を示した．

定理. $\{x_k\} \subseteq \Omega$ となるコンパクト集合 Ω が存在するとき，f に関する適切な仮定のもとで，ある定数 p が存在し少なくとも確率 $1 - 2m\exp(-\frac{\mathcal{C}_0}{4}s) - 2m \cdot 9^s \exp\left(-\frac{\mathcal{C}_1}{4C^6}n\right)$ で次が成り立つ:

$$\min_{k=0,1,\ldots,m-1} \|g_k\|_2 \leq \sqrt{\frac{f(x_0) - \min_{x\in\Omega} f(x)}{mp}}.$$

この定理から $\|g_k\|_2 \leq \varepsilon$ となるために必要な反復回数が高々 $O(\varepsilon^{-2})$ であることがわかるが，この大域的反復計算量は全空間での正則化ニュートン法 [4] で与えられているものと同じである．

参考文献

[1] Claudia D'Ambrosio, Leo Liberti, Pierre-Louis Poirion, and Ky Vu. Random projections for quadratic programs. *Math. Program.*, 183:619–647, 2020.

[2] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[3] Robert Gower, Dmitry Kovalev, Felix Lieder, and Peter Richtárik. RSN: Randomized Subspace Newton. *Adv. Neural Inf. Process. Syst.*, 32:616–625, 2019.

[4] Kenji Ueda and Nobuo Yamashita. Convergence properties of the regularized newton method for the unconstrained nonconvex optimization. *Appl. Math. Optim.*, 62(1):27–46, 2010.