日本オペレーションズ・リサーチ学会
2021年 春季研究発表会

ランダム射影による精度保証付き非凸2次最適化緩和法

申請中　東京大学　*富士晃成　FUJI Terunari
理化学研究所　Pierre-Louis Poirion
01308490　東京大学/理化学研究所　武田朗子　TAKEDA Akiko

1. はじめに

本研究では以下の非凸 2 次最適化問題を近似的に解くことを考える．

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

ここで $Q \in \mathbb{R}^{n\times n}, c \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ であり，$\mathrm{tr}\, Q > 0$ を仮定する．また中心 0，半径 r の閉球 $B(0, r)$ が実行可能領域 $\{x \mid Ax \leq b\}$ に含まれるとする．より一般に $B(x_0, r) \subset \{x \mid Ax \leq b\}$ であるときも変数変換 $y = x - x_0$ を考えることにより先の場合に帰着される．

Johnson-Lindenstrauss の補題 [1] はベクトル空間内の点の集合をそれらの間の距離をほぼ保存したまま低次元空間へ射影できることを示しており，実際にランダム射影が十分に高い確率でこの性質を持つことが知られている [2]．ランダム射影は通常数値データの圧縮に用いられるが，[3] では 2 次最適化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帰着できることが示された．本研究では，非凸 2 次最適化問題をランダム射影により次元削減するとその非凸性が減少することに注目し，非凸 2 次最適化問題を凸かつよりサイズの小さい問題に帰着して近似的に解くことを提案する．またその誤差に関する理論的な評価も与える．

2. ランダム射影

d を n より小さい正整数として各成分 P_{ij} が正規分布 $\mathrm{N}(0, 1/d)$ に独立に従う行列 $P \in \mathbb{R}^{d\times n}$ をランダム行列と呼び，この行列に対応する $\mathbb{R}^n$ から $\mathbb{R}^d$ への写像 $f(x) = Px$ をランダム射影と呼ぶ．このとき，ある定数 $\mathcal{C}_0$ が存在して任意の $x \in \mathbb{R}^n$ と $\varepsilon > 0$ に対して，少なくとも確率 $1-2\exp(-\mathcal{C}_0\varepsilon^2 d)$ で

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

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が大きく変化しないように低次元へ埋め込むことができる． さらにこの性質をもとにして，任意の $x, y \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, Q \in \mathbb{R}^{n\times n}$ について十分高い確率で $x^\top y \approx x^\top P^\top Py, Ax \approx AP^\top Px, x^\top Qx \approx x^\top P^\top PQP^\top Px$ が成り立つことがわかる．すなわちランダム射影はノルムのみならず，内積といった幾何的量も保存する．あるいは $P^\top P$ が単位行列のように振舞うとも解釈できる．

3. ランダム射影による問題

本節では先行研究 [3] の結果について述べる．ランダム行列 P が与えられた時に $\mathbf{P}$ に関連する次の問題を考える．

$$\min_{x \in \mathbb{R}^n} \{x^\top P^\top PQP^\top Px + c^\top P^\top Px \mid AP^\top Px \leq b\}.$$

前節で述べた事実からこの問題は近似的に $\mathbf{P}$ と同じ構造を持っており，それぞれの最適値も近いことが期待される．とくに $u = Px \in \mathbb{R}^d$ とおいて $\bar{Q} = PQP^\top, \bar{c} = Pc, \bar{A} = AP^\top$ を用いれば

$$\mathbf{RP} \equiv \min_{u \in \mathbb{R}^d} \{u^\top \bar{Q}u + \bar{c}^\top u \mid \bar{A}u \leq b\}$$

という射影された問題に帰着され，各最適化問題の最適値を $\mathrm{opt}(\cdot)$ で表すことにすれば $\mathrm{opt}(\mathbf{P}) \approx \mathrm{opt}(\mathbf{RP})$ となることが示されている．また問題のサイズに関しては $d = O(\log(n+m))$ というオーダーが与えられている．

4. ランダム射影の凸化作用と RP の凸緩和

上で述べたように，よりサイズの小さい問題 $\mathbf{RP}$ を解けば近似的に元の問題 $\mathbf{P}$ の最適値が得られることが示されたが，一般には $\mathbf{RP}$ も非凸 2 次最適化問題であり，$\mathbf{RP}$ を解くのにも困難を伴うことが多い．そこで問題の凸性を決定する 2 次の係数行列 $\bar{Q}$ の固有値分布について考察する．簡単な計算から

$$\mathbb{E}[\bar{Q}] = \frac{\mathrm{tr}\, Q}{d} I_d$$

がわかるので $\bar{Q}$ の固有値は $\mathrm{tr}\, Q/d$ 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，$\bar{Q}$ の固有値の多くは正の値をとり，$\bar{Q}$ が近似的に半正定値行列となる．言い方を換えれば，問題 $\mathbf{P}$ が"平均的に凸"であれば問題 $\mathbf{RP}$ は"ほとんど凸"となる．

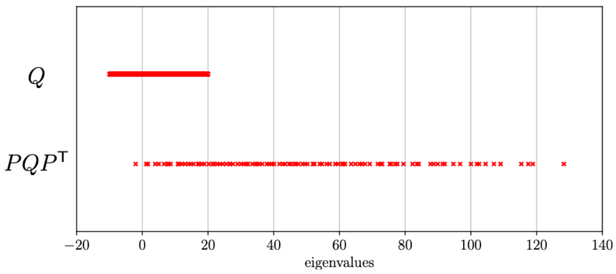

図 1: Q および PQP^T の固有値分布の一例 ($Q \in \mathbb{R}^{1000\times1000}, P \in \mathbb{R}^{200\times1000}$).

実際にある Q に対して $\bar{Q}=PQP^\mathsf{T}$ を計算した時のそれぞれの固有値の分布を図 1 に示す．Q は負の固有値が多く持つが，$\bar{Q}=PQP^\mathsf{T}$ には負の固有値がほとんど存在しないことが確認できる．

このように，$\mathbf{RP}$ の 2 次の係数行列 $\bar{Q}$ のわずかな負のスペクトルを無視することにより構築される行列 $\bar{Q}$ の半正定値錐への射影を $\mathcal{F}^+(\bar{Q})$ と書けば

$$\bar{Q}^+ \equiv \mathcal{F}^+(\bar{Q}) \approx \bar{Q}$$

が成り立つことが期待される．よって，$\mathbf{RP}$ を凸緩和した次の問題

$$\mathbf{CRP} \equiv \min_{u\in\mathbb{R}^d}\{u^\mathsf{T}\bar{Q}^+u + \bar{c}^\mathsf{T}u \mid \bar{A}u \le b\}$$

を考えれば $\mathrm{opt}(\mathbf{CRP}) \approx \mathrm{opt}(\mathbf{RP})$ となる．

問題 $\mathbf{CRP}$ は $\mathbf{P}$ よりサイズが小さく，さらに凸最適化問題なので $\mathbf{P}$ と比べてはるかに解きやすい．また先行研究 [3] の結果と合わせれば $\mathrm{opt}(\mathbf{CRP}) \approx \mathrm{opt}(\mathbf{P})$ となるので，$\mathbf{P}$ の代わりに $\mathbf{CRP}$ を解けば高確率で最適値の近似値が得られる．

最後に $\mathrm{opt}(\mathbf{CRP}) \approx \mathrm{opt}(\mathbf{P})$ がどのような意味で成り立つのか，数学的な主張を述べる．

定理.

問題 $\mathbf{CRP}$ は問題 $\mathbf{P}$ の緩和になっており，

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

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，d 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．

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

ここで x^* を問題 $\mathbf{P}$ 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 $(x^*{x^*}^\mathsf{T}, x^*), (Q, c)$ 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．

$\varepsilon, \varepsilon_1, \varepsilon_2, \varepsilon_3$ がそれぞれ十分に 0 に近ければ

$$\alpha \approx 1,\ \ \frac{3\varepsilon_1+2\varepsilon_2+\varepsilon_3}{\cos\theta^*} \approx 0$$

となるので，上の定理は高確率で $\mathrm{opt}(\mathbf{CRP}) \approx \mathrm{opt}(\mathbf{P})$ が成り立つことを保証している．

定理中の d のみたすべき条件あるいは数値実験については発表中に述べる．

参考文献

[1] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In G. Hedlund, editor, *Conference in Modern Analysis and Probability*, volume 26 of *Contemporary Mathematics*, pages 189–206, Providence, 1984. American Mathematical Society.

[2] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

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