Science & Technology Facilities Council

Technical Report

RAL-TR-2019-005

On monotonic estimates of the norm of the minimizers of regularized quadratic functions in Krylov spaces

C Cartis, NIM Gould, M Lange

April 2019

Enquiries concerning this report should be addressed to:

RAL Library
STFC Rutherford Appleton Laboratory
Harwell Oxford
Didcot
OX11 0QX

Tel: +44(0)1235 445384
Fax: +44(0)1235 446403
email: libraryral@stfc.ac.uk

Science and Technology Facilities Council reports are available online at: http://epubs.stfc.ac.uk

ISSN 1358-6254

On monotonic estimates of the norm of the minimizers of regularized quadratic functions in Krylov spaces*

Coralia Cartis†, Nicholas I. M. Gould‡ and Marius Lange§

ABSTRACT

We show that the minimizers of regularized quadratic functions restricted to their natural Krylov spaces increase in Euclidean norm as the spaces expand.

1 Introduction

Given a real symmetric, possibly indefinite, matrix H and vector g, we are concerned with Krylov methods for approximating the global solution of the possibly nonconvex regularization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad Q(x;\sigma,p) := \tfrac{1}{2}x^T Hx + g^T x + \tfrac{1}{p}\sigma\|x\|^p \tag{1.1}$$

where $\sigma > 0$ $p > 2$ and $\|\cdot\|$ is the Euclidean norm— note that Q is bounded below over $\mathbb{R}^n$, and the global minimizer is well defined. Such methods have been advocated by a number of authors, e.g., [1–3, 7]. Here we are interested in how the norms of the estimates of the solution, and the corresponding "multipliers" $\sigma\|x\|^{p-2}$, evolve as the Krylov process proceeds. The main utility is that these estimates provide useful predictions for the multipliers as the Krylov subspace expands [8]. Our result is an analogue of that obtained by Lukšan, Matonoha and Vlček [10] for the trust-region subproblem.

2 The main result

We start with four vital lemmas that we use to prove our main result. The first shows a simple property of the conjugate gradient method. We use the generic notation $B \succeq 0$ (resp. $B \succ 0$) to mean that the real, symmetric matrix B is positive definite (resp. positive semi-definite).

*This work was supported by the EPSRC grant EP/M025179/1 (NIMG), and both by a German research foundation (DFG) fellowship through the Graduate School of Quantitative Biosciences Munich (QBM) and by the Joachim Herz Stiftung (ML).

†Mathematical Institute, Oxford University, Oxford OX2 6GG, England. Email: coralia.cartis@maths.ox.ac.uk

‡Scientific Computing Department, STFC-Rutherford Appleton Laboratory, Chilton OX11 0QX, England. Email: nick.gould@stfc.ac.uk.

§Institute of Computational Biology, Helmholtz Zentrum München—German Research Center for Environmental Health, 85764 Neuherberg, Germany and Department of Mathematics, Technische Universität München, 85748 Munich, Germany. Email: marius.lange@helmholtz-muenchen.de.

Lemma 2.1. Given a real symmetric matrix B and real vector g, let

$$\mathcal{K}_k(B,g) := \text{span}\left\{g, Bg, \ldots, B^{k-1}g\right\},$$

$k \geq 1$, be the k-th Krylov subspace generated by B and the vector g, and let the columns of V_k provide an orthonormal basis for $\mathcal{K}_k(B,g)$. Letting $\ell \geq k \geq 1$, suppose that

$$V_\ell^T B V_\ell \succ 0 \tag{2.1}$$

and define

$$x_k = \underset{x \in \mathcal{K}_k(B,g)}{\arg\min} \quad Q(x) := \tfrac{1}{2} x^T B x + g^T x.$$

Then

$$\|x_k\| \leq \|x_\ell\|.$$

Proof. This follows from [4, Thm.7.5.1] as the requirement there, namely that $p_k^T B p_k > 0$ for specific vectors $p_k \in \mathcal{K}_k(B,g)$, is implied by the more general assumption (2.1). □

Note that this is a generalization of [11, Thm.2.1] that relaxes the requirement that B be everywhere positive definite to be so merely over the evolving Krylov subspaces of interest.

Our second lemma compares Krylov subspaces of the matrices B and $B + \mu I$ for some $\mu \in \mathbb{R}$.

Lemma 2.2. [10, Lem.2.3]. Let B, g and $\mathcal{K}_k$ be as in Lemma 2.1, and $\mu \in \mathbb{R}$. Then

$$\mathcal{K}_k(B + \mu I, g) = \mathcal{K}_k(B,g). \tag{2.2}$$

Next, we state a crucial relation between the values of the Lagrange multipliers and the norms of the direction vectors.

Lemma 2.3. [10, Lem.2.5]. Suppose that the columns of V_k provide an orthonormal basis for $\mathcal{K}_k(H,g)$ for given real symmetric H and real g. Let $V_k^T H V_k + \mu_i I$, $\mu_i \in \mathbb{R}$, $i \in \{1,2\}$, be symmetric and positive definite. Let

$$x_k(\mu_i) = \underset{x \in \mathcal{K}_k(H,g)}{\arg\min} \; Q_{\mu_i}(x) := \tfrac{1}{2} x^T (H + \mu I)x + g^T x.$$

Then

$$\mu_2 \leq \mu_1 \;\text{ if and only if }\; \|x_k(\mu_2)\| \geq \|x_k(\mu_1)\|.$$

Out final lemma indicates that the evolving minimizers are unique.

Lemma 2.4. Let H, g and V_k be as in Lemma 2.3, and let the grade [9] $m \leq n$ be the maximum dimension of the evolving Krylov spaces $\mathcal{K}_k(H,g)$, $k = 1, \ldots, n$. Then $V_k^T H V_k + \mu_k I \succ 0$ for all $1 \leq k \leq m$.

Proof. Using the Lanczos orthonormal basis, we have that $V_k^T H V_k = T_k$ for an irreducible tridiagonal matrix T_k for $k = 1, \ldots, m$. It then follows [4, Thm.7.5.12] that the "hard case" cannot occur, and thus that the only possible root μ_k of the secular equation for the problem [5, Sec.2.2] satisfies $\mu_k > -\lambda_{\min}(T_k)$, where $\lambda_{\min}$ denotes the leftmost eigenvalue of its symmetric matrix argument. □

We are now in a position to state and prove our main theorem.

Theorem 2.5. Given a real symmetric matrix H, vector g and scalars $\sigma > 0$ and $p > 2$, let

$$x_j = \underset{x \in \mathcal{K}_j(H,g)}{\arg\min} \; Q(x; \sigma, p) := \tfrac{1}{2} x^T H x + g^T x + \tfrac{1}{p} \sigma \|x\|^p,$$

and let

$$\mu_j = \sigma \|x_j\|^{p-2} \tag{2.3}$$

for $j \geq 1$. Then $\mu_k \leq \mu_\ell$ and $\|x_k\| \leq \|x_\ell\|$ for $1 \leq k \leq \ell \leq m$.

Proof. Let V_j be as in the statement of Lemma 2.3. The vector $x_j = V_j y_j$ is a minimizer of the j-th regularization subproblem if and only if

$$V_j^T (H + \mu_j I) V_j y_j = -V_j^T g, \;\; V_j^T (H + \mu_j I) V_j \succeq 0, \;\text{ and }\; \mu_j = \sigma \|y_j\|^{p-2}, \tag{2.4}$$

and the minimizer is unique since $V_j^T (H + \mu_j I) V_j \succ 0$ from Lemma 2.4 [5, Thm.2].

Consider two integers k and ℓ for which $1 \le k \le \ell \le m$.

Since we have $V_k^T(H + \mu_k I)V_k \succ 0$ and $V_\ell^T(H + \mu_\ell I)V_\ell \succ 0$, and as $\mathcal{K}_k(H + \mu_k I, g) = \mathcal{K}_k(H, g)$ by Lemma 2.2, it follows from (2.4) that x_k is also the (unique) solution of the constrained minimization problem

$$x_k = \underset{x \in \mathcal{K}_k(H,g)}{\arg\min} \quad Q_{\mu_k}(x), \quad \text{where} \quad Q_\mu(x) = \tfrac{1}{2} x^T (H + \mu I)x + g^T x.$$

Assume that $\mu_k > \mu_\ell$, which implies that $V_\ell^T(H + \mu_k I)V_\ell \succ 0$. Let

$$x_\ell(\mu_k) = \underset{x \in \mathcal{K}_\ell(H,g)}{\arg\min} \quad Q_{\mu_k}(x).$$

Then it follows from Lemma 2.1 that

$$\|x_k\| \le \|x_\ell(\mu_k)\|. \tag{2.5}$$

But since $\mu_\ell < \mu_k$, Lemma 2.3 gives that

$$\|x_\ell(\mu_k)\| \le \|x_\ell(\mu_\ell)\| = \|x_\ell\|. \tag{2.6}$$

Hence using the definition (2.3) and combining the inequalities (2.5) and (2.6)

$$\mu_k = \sigma \|x_k\|^{p-2} \le \sigma \|x_\ell\|^{p-2} = \mu_\ell < \mu_k$$

which is a contradiction. Thus $\mu_k \le \mu_\ell$ has to hold. It then follows from the definition (2.3) that $\|x_k\| \le \|x_\ell\|$. □

The monotonic behaviour of the multipliers μ_k was predicted in [8, Lem.2.6] when $p = 3$, but the proof suggested there relied on [10, Thm.2.6], which appears to have a minor flaw—the proof depends on [11, Thm.2.1], but applies this at one point to an indefinite $H + \mu I$. Lemma 2.1 avoids this issue, and the same result fixes the proof of [10, Thm.2.6].

3 Comments and conclusions

We have shown that the norms of the approximations generated by well-known Krylov methods for solving the regularization problem (1.1) increase monotonically as the dimension of the Krylov spaces expands. This implies that the corresponding "multipliers" also increase, and is useful as estimates of these multipliers are crucial when solving the Krylov subproblem; in particular, as the multiplier for the k-th problem is a lower bound for the $k{+}1$-st, Newton-like iterations will converge both globally and rapidly to μ_{k+1} when started from μ_k if additionally $\mu_k > \lambda_{\min}(T_{k+1})$. [5, §3]. Knowledge of the monotonic nature of these quantities is also important when deriving convergence bounds [6] for such methods.

We warn readers that in exceptional circumstances, namely that g is orthogonal to the eigenspace corresponding to the leftmost eigenvalue of H and σ is insufficiently large, the global minimizer of (1.1) will not lie in $\mathcal{K}_m(H,g)$, and μ_m will underestimate the optimal multiplier. This (zero-probability) possibility is often referred to as the "hard case", [3, §6,1], and, despite their popularity, might be viewed as an unavoidable defect of Krylov methods.

The main result here may trivially be extended for Krylov methods to

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; Q(x;\sigma,p,M) := \tfrac{1}{2}x^T H x + g^T x + \tfrac{1}{p}\sigma \|x\|_M^p,$$

for given symmetric $M \succ 0$, where $\|x\|_M^2 := x^T M x$, so long as we instead consider the Krylov spaces $\mathcal{K}(M^{-1}H, M^{-1}g)$. It is well known that this may be achieved using the M-preconditioned Lanczos method [3, Sec.6.3]. In particular, if

$$x_j = \underset{x \in \mathcal{K}_j(M^{-1}H,M^{-1}g)}{\arg\min} \;\; Q(x;\sigma,p,M) \;\text{ and }\; \mu_j = \sigma\|x_j\|_M^{p-2},$$

it follows (using the transformation $x \leftarrow M^{\frac{1}{2}}x$) that

$$\mu_k \le \mu_\ell \;\text{ and }\; \|x_k\|_M \le \|x_\ell\|_M$$

for $1 \le k \le \ell \le m$ just as in Theorem 2.5.

References

[1] T. Bianconcini, G. Liuzzi, B. Morini, and M. Sciandrone. On the use of iterative methods in cubic regularization for unconstrained optimization. Rapporto Tecnico 6/2013, Dipartimento di Ingegneria Industriale, Universita degli Studi di Firenze, Italy, 2013.

[2] Y. Carmon and J. C. Duchi. Analysis of Krylov subspace solutions of regularized nonconvex quadratic problems. arXiv:1806.09222v1, 2018.

[3] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming, Series A*, 127(2):245–295, 2011.

[4] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, 2000.

[5] N. I. M. Gould, D. P. Robinson, and H. S. Thorne. On solving trust-region and other regularised subproblems in optimization. *Mathematical Programming Computation*, 2(1):21–57, 2010.

[6] N. I. M. Gould and V. Simoncini. Error estimates for iterative algorithms for minimizing regularized quadratic subproblems. Technical Report RAL-TR-2019-004, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2019.

[7] J. M. Kohler and A. Lucchi. Sub-sampled cubic regularization for non-convex optimization. arxiv.org/abs/1705.05933v3, 2017.

[8] M. Lange. Subproblem solutions in cubic regularisation methods. M.Sc. thesis, University of Oxford, England, 2017.

[9] J. Liesen and Z. Strakoš. *Krylov subspace methods*. Oxford University Press, Oxford, 2013.

[10] L. Lukšan, C. Matonoha, and J. Vlček. On Lagrange multipliers of trust-region subproblems. *BIT*, 48(4):763–768, 2008.

[11] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis*, 20(3):626–637, 1983.