Strong local convergence properties of adaptive regularized methods for nonlinear least-squares

STRONG LOCAL CONVERGENCE PROPERTIES OF ADAPTIVE REGULARIZED METHODS FOR NONLINEAR LEAST-SQUARES *

S. BELLAVIA† AND B. MORINI†

Abstract. This paper studies adaptive regularized methods for nonlinear least-squares problems where the model of the objective function used at each iteration is either the Euclidean residual regularized by a quadratic term or the Gauss-Newton model regularized by a cubic term. For suitable choices of the regularization parameter the role of the regularization term is to provide global convergence. In this paper we investigate the impact of the regularization term on the local convergence rate of the methods and establish that, under the well-known error bound condition, quadratic convergence to zero-residual solutions is enforced. This result extends the existing analysis on the local convergence properties of adaptive regularized methods. In fact, the known results were derived under the standard full rank condition on the Jacobian at a zero-residual solution while the error bound condition is weaker than the full rank condition and allows the solution set to be locally nonunique.

Keywords: Nonlinear least-squares problems, regularized models, error bound condition, local convergence.

1. Introduction. In this paper we discuss the local convergence behaviour of adaptive regularized methods for solving the nonlinear least-squares problem

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\|F(x)\|_2^2, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a given vector-valued continuously-differentiable function.

Adaptive regularization approaches for unconstrained optimization base each iteration upon a quadratic or a cubic regularization of standard models for f. Their construction follows from observing that, for suitable choices of the regularization parameter, the regularized model overestimates the objective function. The role of the adaptive regularization is to control the distance between successive iterates and to provide global convergence of the procedures. Interesting connections among these approaches, trust-region and linesearch methods are established in [23].

First ideas on adaptive cubic regularization of the Newton's quadratic model for f can be found in [15] in the context of affine invariant Newton methods; further results were successively obtained in [25]. In [21] it was shown that the use of a local cubic overestimator for f yields an algorithm with global and fast local convergence properties and a good global iteration complexity. Elaborating on these ideas, an adaptive cubic regularization method for unconstrained minimization problems was proposed in [4]; it employs a cubic regularization of the Newton's model and allows approximate minimization of the model and/or a symmetric approximation to the Hessian matrix of f. This new approach enjoys good global and local convergence properties as well as the same (in order) worst-case iteration complexity bound as in [21].

Adaptive regularized methods have been also studied for the specific case of nonlinear least-square problems. Complexity bounds for the method proposed in [4] applied to potentially rank-deficient nonlinear least-squares problems were given in [8]. Moreover, such an approach was specialized to the solution of (1.1) along with

†Dipartimento di Ingegneria Industriale, Università di Firenze, viale G.B. Morgagni 40, 50134 Firenze, Italia, stefania.bellavia@unifi.it, benedetta.morini@unifi.it

* Work partially supported by INdAM-GNCS, under the 2013 Projects *Numerical Methods and Software for Large-Scale Optimization with Applications to Image Processing.*

the updating rules for the regularization parameter [14]. Regarding quadratic regularization, a model consisting of the Euclidean residual regularized by a quadratic term was proposed in [20] for nonlinear systems and then extended to general nonlinear least-squares problems allowing the use of approximate minimizers of the model [2].

Further recent works on adaptive cubic regularization concern its extension to constrained nonlinear programming problems [7, 8] and its application in a barrier method for solving a nonlinear programming problem [3].

In this paper we focus on two adaptive regularized methods for nonlinear least-square problems introduced in [2, 14] and investigate their local convergence properties. The model used in [2] is a Euclidean residual regularized by a quadratic term, whereas the model used in [14] is the Gauss-Newton model regularized by a cubic term. These methods are especially suited for computing zero-residual solutions of (1.1) which is a case of interest for example when F is the map of a square nonlinear system of equations or it models the detection of feasible points in nonlinear programming [11, 18]

The two procedures considered are known to be quadratically convergent to zero-residual solutions where the Jacobian of F is full rank, see [2, 4]. Here we go a step further and show that the presence of the regularization term provides fast local convergence under weaker assumptions. Thus we can conclude that the regularization has a double role in enhancing the properties of the underlying unregularized methods: besides guaranteeing global convergence, it enforces strong local convergence properties.

Our local convergence analysis concerns zero-residual solutions of (1.1) satisfying an error bound condition and covers square problems ($m = n$), overdetermined problems ($m > n$) and underdetermined problems ($m < n$). Error bounds were introduced in mathematical programming in order to bound, in terms of a computable residual function, the distance of a point from the typically unknown solution set [22]. These conditions have been widely used for studying the local convergence of various optimization methods: proximal methods for minimization problems [16], Newton-type methods for complementarity problems [24], derivative-free methods for nonlinear least-squares [27], Levenberg-Marquardt methods for constrained and unconstrained nonlinear least-squares [1, 10, 12, 13, 17, 18, 26, 28]. Our study is motivated by these latter results and the connection established here between the Levenberg-Marquadt methods and the adaptive regularized methods. A similar insight on such a connection between the two classes of procedures is given in [3].

Following [26] we consider the case where the norm of F provides a local error bound for (1.1) and prove Q-quadratic convergence of ARQ and ARC methods. The error bound condition considered can be valid at locally nonunique zero-residual solutions of (1.1). In fact, it may hold, irrespective of the dimensions m and n of F, at solutions where the Jacobian J of F is not full rank. Thus our study establishes novel local convergence properties of ARQ and ARC under milder conditions than the standard full rank condition on J assumed in [2, 4]. In the context of Levenberg-Marquardt methods, such fast local convergence properties were defined as *strong* properties, see [17].

The paper is organized as follows: in §2 we describe the two methods under study, discuss their connection with the Levenberg-Marquardt methods and the error bound condition used. In §3 we provide a convergence result that paves the way for the local convergence analysis carried out in §4. In this latter section we also show how to compute an approximate minimizer of the model and retain fast local convergence

properties of the studied approaches. In §5 we discuss the case where the approximate step is computed by minimizing the model in a subspace and its implication on the local convergence properties. In §6 we make some concluding remarks.

Notations. For the differentiable mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, the Jacobian matrix of F at x is denoted by $J(x)$. The gradient and the Hessian matrix of the smooth function $f(x) = \|F(x)\|^2/2$ are denoted by $g(x) = J(x)^T F(x)$ and $H(x)$ respectively. When clear from the context, the argument of a mapping is omitted and, for any function h, the notation h_k is used to denote $h(x_k)$. The 2-norm is denoted by $\|x\|$. For any vector $y \in \mathbb{R}^n$, the ball with center y and radius ρ is indicated by $B_\rho(y)$, i.e. $B_\rho(y) = \{x : \|x - y\| \leq \rho\}$. The identity matrix $n \times n$ is indicated by I.

2. Adaptive regularized quadratic and cubic algorithms. In this section we discuss the adaptive quadratic and cubic regularized methods proposed in [2, 4, 14] and summarize some of their properties.

We first introduce the models proposed for problem (1.1). Given some iterate x_k, in [2] a model consisting of the euclidean residual $\|F_k + J_k p\|$ regularized by a quadratic term is introduced. Letting σ_k be a dynamic strictly positive parameter, the model takes the form

$$m_k^Q(p) = \|F_k + J_k p\| + \sigma_k \|p\|^2. \tag{2.1}$$

If the Jacobian J of F is globally Lipschitz continuous (with constant $2L$) and $\sigma_k = L$, then m_k^Q reduces to the modified Gauss-Newton model proposed in [20]. Whenever $\sigma_k \geq L$, $\|F\|$ is overestimated around x_k by means of m_k^Q , i.e.

$$\|F(x_k + p)\| \leq m_k^Q(p).$$

Alternatively, in [14] the cubic regularization of the Gauss-Newton model

$$m_k^C(p) = \frac{1}{2}\|F_k + J_k p\|^2 + \frac{1}{3}\sigma_k \|p\|^3, \tag{2.2}$$

is used with a dynamic strictly positive parameter σ_k. This model is motivated by the cubic overestimation of f devised in [4, 15, 21]. In fact, if the Hessian H of f is Lipschitz continuous (with constant $2L$), then

$$f(x_k + p) \leq f_k + p^T g_k + \frac{1}{2} p^T H_k p + \frac{1}{3} L ||p||^3.$$

Thus, (2.2) is obtained replacing L by σ_k and considering the first order approximation $J_k^T J_k$ to H_k; it is well-known that the latter approximation is reasonable in a neighborhood of a zero-residual solution of problem (1.1) [11].

Before addressing the use of the models m_k^Q and m_k^C in the solution of (1.1), we review their properties and the form of the minimizer.

LEMMA 2.1. *Suppose that $\sigma_k > 0$. Then the models m_k^Q and m_k^C are strictly convex.*

Proof. The strict convexity of m_k^Q is proved in [2, Lemma 2.1]. Regarding m_k^C, the function $\|F_k + J_k p\|^2$ is convex and the function $\|p\|^3$ is strictly convex. □

LEMMA 2.2. *Let $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ be continuously differentiable and suppose $\|g_k\| \neq 0$. Then,*

i) If p_k^* is the minimizer of m_k^Q, then there is a nonnegative λ_k^* such that (p_k^*, λ_k^*) solves

$$(J_k^T J_k + \lambda I)p = -g_k, \tag{2.3}$$

$$\lambda = 2\sigma_k \|J_k p + F_k\|. \tag{2.4}$$

Moreover, λ_k^* is such that

$$\lambda_k^* \in [0, 2\sigma_k \|F_k\|\,]. \tag{2.5}$$

ii) If there exists a solution (p_k^*, λ_k^*) of (2.3) and (2.4) with $\lambda_k^* > 0$, then p_k^* is the minimizer of m_k^Q. Otherwise, the minimizer of m_k^Q is given by the minimum norm solution of the linear system $J_k^T J_k p = -g_k$.

iii) The vector p_k^* is the unique minimizer of m_k^C if and only if there exists a positive λ_k^* such that (p_k^*, λ_k^*) solves (2.3) and

$$\lambda_k^* = \sigma_k \|p_k^*\|. \tag{2.6}$$

Proof. The results for m_k^Q are given in [2, Lemma 4.1, Lemma 4.3]. The hypothesis for m_k^C follows from the application of [4, Theorem 3.1] to such a model. □

The above lemma shows that the minimizer of the quadratic and cubic regularized models solves the shifted linear system (2.3) for a specific value $\lambda = \lambda_k^*$ which depends on the model and is given in (2.4) and (2.6) respectively.

In the rest of the paper, the notation m_k will be used to indicate the model, irrespective of its specific form, in all the expressions that are valid for both m_k^Q and m_k^C. Further, for a given $\lambda \geq 0$, we let $p(\lambda)$ be the minimum-norm solution of (2.3), and $p_k^* = p(\lambda_k^*)$ be the minimizer in Lemma 2.2 without distinguishing between the models; it will be inferred from the context whether p_k^* minimizes either m_k^Q or m_k^C. The vector $p(\lambda)$ can be characterized in terms of the singular values of J_k as follows.

LEMMA 2.3. *[2, Lemma 4.2] Assume $\|g_k\| \neq 0$ and let $p(\lambda)$ be the minimum norm solution of (2.3) with $\lambda \geq 0$. Assume furthermore that J_k is of rank ℓ and its singular-value decomposition is given by $U_k \Sigma_k V_k^T$ where $\Sigma_k = \operatorname{diag}(\varsigma_1^k, \ldots, \varsigma_\nu^k)$, with $\nu = \min(m, n)$. Then, denoting $r_k = ((r_k)_1, (r_k)_2, \ldots, (r_k)_\nu)^T = U_k^T F_k$, we have that*

$$\|p(\lambda)\|^2 = \sum_{i=1}^{\ell} \frac{(\varsigma_i^k (r_k)_i)^2}{((\varsigma_i^k)^2 + \lambda)^2}. \tag{2.7}$$

In the literature, adaptive regularized methods for (1.1) compute the step from one iterate to the next as an approximate minimizer of either m_k^Q or m_k^C and test its progress toward a solution. In [2] a procedure based on the use of the model m_k^Q for all $k \geq 0$ is proposed; here it is named Adaptive Quadratic Regularization (ARQ). On the other hand, in [14] a procedure based on the use of the model m_k^C for all $k \geq 0$ is given and denoted Adaptive Cubic Regularization (ARC).

The description of kth iteration of ARQ and ARC is summarized in Algorithm 2.1 where the string METHOD denotes the name of the method, i.e. it is either 'ARQ' or 'ARC'. The trial step selection consists in finding an approximate minimizer p_k of the model which produces a value of m_k smaller than that achieved by the Cauchy point (2.8). This step is accepted and the new iterate x_{k+1} is set to $x_k + p_k$ if a sufficient decrease in the objective is achieved; otherwise, the step is rejected and x_{k+1} is set to

x_k. We note that the denominator in (2.10) and (2.11) is strictly positive whenever the current iterate is not a first-order critical point. As a consequence, the algorithm is well defined and the sequence $\{f_k\}$ is non-increasing. The rules for updating the parameter σ_k parallel those for updating the trust-region size in trust-region methods [2, 4].

Algorithm 2.1 and standard trust-region methods [9] belong to the same unifying framework, see [23]. In a trust-region method with Gauss-Newton model the step is an approximate minimizer of the model subject to the explicit constraint $\|p_k\| \leq \Delta_k$ for some adaptive trust-region radius Δ_k. In the adaptive regularized methods, p_k is an approximate minimizer of a regularized Gauss-Newton model and the stepsize is implicitly controlled.

Algorithm 2.1: kth iteration of ARQ and ARC

Given x_k and the constants $\sigma_k > 0$, $1 > \eta_2 > \eta_1 > 0$, $\gamma_2 \geq \gamma_1 > 1$, $\gamma_3 > 0$.
If METHOD='ARQ' let m_k be m_k^Q, else let m_k be m_k^C.

Step 1: Set

$$p_k^c = -\alpha_k g_k, \qquad \alpha_k = \operatorname*{argmin}_{\alpha \geq 0} m_k(-\alpha g_k). \tag{2.8}$$

Compute an approximate minimizer p_k of $m_k(p)$ such that

$$m_k(p_k) \leq m_k(p_k^c). \tag{2.9}$$

Step 2: If METHOD='ARQ' compute

$$\rho_k = \frac{\|F(x_k)\| - \|F(x_k + p_k)\|}{\|F(x_k)\| - m_k^Q(p_k)}, \tag{2.10}$$

else compute

$$\rho_k = \frac{\frac{1}{2}\|F(x_k)\|^2 - \frac{1}{2}\|F(x_k + p_k)\|^2}{\frac{1}{2}\|F(x_k)\|^2 - m_k^C(p_k)}. \tag{2.11}$$

Step 3: Set

$$x_{k+1} = \begin{cases} x_k + p_k & \text{if } \rho_k \geq \eta_1, \\ x_k & \text{otherwise.} \end{cases}$$

Step 4: Set

$$\sigma_{k+1} \in \begin{cases} (0, \sigma_k] & \text{if } \rho_k \geq \eta_2 & \text{(very successful)}, \\ [\sigma_k, \gamma_1 \sigma_k\,) & \text{if } \eta_1 \leq \rho_k < \eta_2 & \text{(successful)}, \\ [\gamma_1 \sigma_k, \gamma_2 \sigma_k\,) & \text{otherwise} & \text{(unsuccessful)}. \end{cases} \tag{2.12}$$

The convergence properties of ARQ and ARC methods have been studied in [2] and [4] under standard assumptions for unconstrained nonlinear least-squares problems and optimization problems respectively. Both ARQ and ARC show global convergence to first-order critical points of (1.1), see [2, Theorem 3.8], [4, Corollary 2.6].

Moreover, imposing a certain level of accuracy in the computation of the approximate minimizer p_k, quadratic asymptotic convergence to a zero-residual solution is achieved. Specifically, the sequence generated $\{x_k\}$ is Q-quadratically convergent to a zero-residual solution x^* if $J(x^*)$ is full rank; this result is valid for ARQ under any relation between the dimensions m and n of F ([2, Theorems 4.9, 4.10, 4.11]) while it is proved for ARC method whenever $m \geq n$ [4, Corollary 4.10].

The purpose of this paper is to show that, even if J is not of full-rank at the solution, ARQ and ARC are Q-quadratically convergent to a zero-residual solution provided that the so-called error bound condition holds. This property is suggested by two issues discussed in the next section: the connection between adaptive regularization methods and the Levenberg-Marquardt method and the local convergence properties of the latter under the error bound condition.

2.1. Connection of the steps in ARQ, ARC and Levenberg-Marquardt methods. In a Levenberg-Marquardt method [19], given x_k and a positive scalar μ_k, the quadratic model for f around x_k takes the form

$$m_k^{LM}(s) = \frac{1}{2}\|F_k + J_k s\|_2^2 + \frac{1}{2}\mu_k \|s\|^2. \tag{2.13}$$

Letting s_k be the minimizer, the new iterate x_{k+1} is set to $x_k + s_k$ if it provides a sufficient decrease of f; otherwise x_{k+1} is set to x_k.

Clearly, the minimizer s_k of m_k^{LM} is the solution of the shifted linear system (2.3) where $\lambda = \mu_k$. The difference between s_k and the step $p(\lambda)$ in ARQ and ARC methods lies in the shift parameter used. In the Levenberg-Marquardt approach the regularization parameter can be chosen as proposed by Moré in the renowned paper [19]. In the adaptive regularized methods, the optimal value of λ_k^* for the minimizer $p_k^* = p(\lambda_k^*)$ depends on the regularization parameter σ_k and satisfies (2.4) in ARQ and (2.6) in ARC. Moreover, for an approximate minimizer $p_k = p(\lambda_k)$ the value of λ_k will be close to λ_k^* on the base of specified accuracy requirements.

In [26], Yamashita and Fukushima showed that Levenberg-Marquardt methods may converge locally quadratically to a zero-residual solutions of (1.1) satisfying a certain error bound condition. Letting $\mathcal{S}$ denote the nonempty set of zero-residual solutions of (1.1) and $d(x, \mathcal{S})$ denote the distance between the point x and the set $\mathcal{S}$, such a condition is defined as follows.

ASSUMPTION 2.1. *A point $x^* \in \mathcal{S}$ satisfies the error bound condition if there exist positive constants χ and α such that*

$$\frac{1}{\alpha}\, d(x, \mathcal{S}) \leq \|F(x)\|, \quad \textit{for all } x \in B_\chi(x^*). \tag{2.14}$$

By extending the theory in [26], Fan and Yuan [13] and Behling and Fischer [1] showed that, under the same condition, Levenberg-Marquardt methods converge locally Q-quadratically provided that

$$\mu_k = O(\|F_k\|^\delta), \quad \text{for some } \delta \in [1, 2]. \tag{2.15}$$

In [17] this property was defined as a *strong* local convergence property since it is weaker than the standard full rank condition on the Jacobian J of F. Inequality (2.14) bounds the distance of vectors in a neighbourhood of x^* to the set $\mathcal{S}$ in terms of the computable residual $\|F\|$ and depends on the solution x^*. Remarkably, it allows the solution set $\mathcal{S}$ to be locally nonunique [17]. Specifically, in case of overdetermined

or square residual functions F, i.e. $m \geq n$, the condition $J(x^*)$ is full rank implies that (2.14) holds, see e.g. [18, Lemma 4.2]. On the other hand, the converse is not true and (2.14) may hold even in the case where $J(x^*)$ is not full rank, irrespective to the relationship between m and n. To see this, consider the example given in [10, p. 608] where $F : \mathbb{R}^2 \to \mathbb{R}^2$ has the form

$$F(x_1, x_2) = \left(e^{x_1 - x_2} - 1,\, (x_1 - x_2)(x_1 - x_2 - 2)\right)^T,$$

We have that $\mathcal{S} = \{x \in \mathbb{R}^2 : x_1 = x_2\}$, J is singular at any point in $\mathcal{S}$. As $d(x, \mathcal{S}) = (\sqrt{2}/2)|x_1 - x_2|$, the error bound condition is satisfied with $\alpha = 1$ in a proper neighbourhood of any point $x^* \in \mathcal{S}$.

Slight modifications of the previous example show that (2.14) is weaker than the full rank condition in the overdetermined and underdetermined cases too. In the first case, an example is given by $F : \mathbb{R}^2 \to \mathbb{R}^3$ such that

$$F(x_1, x_2) = \left(e^{x_1 - x_2} - 1,\, (x_1 - x_2)(x_1 - x_2 - 2),\, sin(x_1 - x_2)\right)^T,$$

The error bound condition can be showed proceeding as before, by noting that $\mathcal{S} = \{x \in \mathbb{R}^2 : x_1 = x_2\}$, J is not full rank at any point in $\mathcal{S}$, $d(x, \mathcal{S}) = (\sqrt{2}/2)|x_1 - x_2|$ and the error bound condition is satisfied with $\alpha = 1$ in a proper neighbourhood of any point $x^* \in \mathcal{S}$.

For the underdetermined case, let us consider the problem where $F : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

$$F(x_1, x_2, x_3) = \left(e^{x_1 - x_2 - x_3} - 1,\, (x_1 - x_2 - x_3)(x_1 - x_2 - x_3 - 2)\right)^T.$$

Then $\mathcal{S} = \{x \in \mathbb{R}^3 : x_1 - x_2 - x_3 = 0\}$ and J is rank deficient everywhere in $\mathcal{S}$. As $d(x, \mathcal{S}) = (\sqrt{3}/3)|x_1 - x_2 - x_3|$, again the error bound condition is satisfied with $\alpha = 1$ in a proper neighbourhood of any point $x^* \in \mathcal{S}$.

In this paper we show that, under Assumption 2.1, ARQ and ARC exhibit the same strong convergence properties as the Levenberg-Marquardt methods. Since the existing local results in literature are valid as long as $J(x^*)$ is of full-rank, our new results offer a further insight into the effects of the regularizations employed in ARQ and ARC.

3. Local convergence of a sequence. In this section we analyze the local behaviour of a sequence $\{x_k\}$ admitting a limit point x^* in the set $\mathcal{S}$ of the zero-residual solution of (1.1). For x_k sufficiently close to x^* and under suitable conditions on the step taken and the behaviour of $d(x_k, \mathcal{S})$, we show that the sequence $\{x_k\}$ converges to x^* Q-quadratically.

The theorem proved below uses technicalities from [13] but it does not involve the error bound condition. It will be used in §4 because we will show that, under Assumption 2.1, the sequences generated by the two Adaptive Regularized methods satisfy its assumptions.

THEOREM 3.1. *Suppose that $x^* \in \mathcal{S}$ and that $\{x_k\}$ is a sequence with limit point x^*. Let $q_k \in \mathbb{R}^n$ be such that,*

$$\|q_k\| \leq \Psi d(x_k, \mathcal{S}), \quad \text{if } x_k \in B_\epsilon(x^*), \tag{3.1}$$

for some positive Ψ and ϵ, and

$$x_{k+1} = x_k + q_k, \quad \text{if } x_k \in B_\psi(x^*), \tag{3.2}$$

for some $\psi \in (0, \epsilon)$. Then, $\{x_k\}$ converges to x^* Q-quadratically whenever

$$d(x_k + q_k, \mathcal{S}) \leq \Gamma d(x_k, \mathcal{S})^2, \quad \text{if } x_k \in B_\psi(x^*), \tag{3.3}$$

for some positive Γ.

Proof. In order to prove that the sequence $\{x_k\}$ is convergent, we show that it is a Cauchy sequence. Consider a fixed positive scalar ζ such that

$$\zeta \leq \min\left\{\frac{\psi}{1+4\Psi}, \frac{1}{2\Gamma}\right\}. \tag{3.4}$$

We first prove that if $x_k \in B_\zeta(x^*)$, then

$$x_{k+\ell} \in B_\psi(x^*), \tag{3.5}$$

for all $\ell \geq 1$. Consider the case $\ell = 1$ first. Since $x_k \in B_\zeta(x^*)$ and $\zeta < \psi$, it follows from (3.2) that $x_{k+1} = x_k + q_k$ and by (3.1) we get $\|x_k + q_k - x^*\| \leq \|x_k - x^*\| + \|q_k\| \leq \zeta(1+\Psi)$. Then, (3.4) gives $x_k + q_k \in B_\psi(x^*)$ and (3.5) holds for $\ell = 1$. Assume now that (3.5) holds for iterations $k+j$, $j = 0, \ldots, \ell-1$. By (3.1) and (3.2)

$$\begin{aligned}
\|x_{k+\ell} - x^*\| &\leq \|x_{k+\ell} - x_{k+\ell-1}\| + \ldots + \|x_k - x^*\| \\
&\leq \zeta + \sum_{j=0}^{\ell-1} \|q_{k+j}\| \\
&\leq \zeta + \Psi \sum_{j=0}^{\ell-1} d(x_{k+j}, \mathcal{S}). \qquad (3.6)
\end{aligned}$$

To provide an upper bound for (3.6), we use (3.3) and obtain

$$d(x_{k+j}, \mathcal{S}) \leq \Gamma d(x_{k+j-1}, \mathcal{S})^2 \leq \ldots \leq \Gamma^{(2^j-1)} d(x_k, \mathcal{S})^{2^j} \leq \Gamma^{(2^j-1)} \zeta^{2^j}, \tag{3.7}$$

for $j = 0, \ldots, \ell-1$. Moreover, since (3.4) implies $\Gamma\zeta \leq \frac{1}{2}$, it follows

$$d(x_{k+j}, \mathcal{S}) \leq 2\zeta \left(\frac{1}{2}\right)^{2^j}, \tag{3.8}$$

and

$$\sum_{j=0}^{\ell-1} d(x_{k+j}, \mathcal{S}) \leq 2\zeta \sum_{j=0}^{\ell-1} \left(\frac{1}{2}\right)^{2^j}.$$

Thus, since $2^j > j$ for $j \geq 0$, we get

$$\sum_{j=0}^{\ell-1} d(x_{k+j}, \mathcal{S}) \leq 2\zeta \sum_{j=0}^{\ell-1} \left(\frac{1}{2}\right)^{j} \leq 4\zeta,$$

and by (3.6)

$$\|x_{k+\ell} - x^*\| \leq \zeta(1+4\Psi) \leq \psi,$$

where in the last inequality we have used the definition of ζ in (3.4). Then, we have proved (3.5) and by (3.2) we can conclude that $x_{k+j+1} = x_{k+j} + q_{k+j}$, for $j \geq 0$. Using (3.8) and proceeding as above we have

$$\|x_{k+r} - x_{k+t}\| \leq \sum_{j=r}^{t-1} \|q_{k+j}\| \leq \Psi \sum_{j=r}^{t-1} d(x_{k+j}, \mathcal{S}) \leq 4\Psi\zeta.$$

Then $\{x_k\}$ is a Cauchy sequence and it is convergent. Since x^* is a limit point of the sequence we deduce that $x_k \to x^*$.

We finally show the convergence rate of the sequence. Let k sufficiently large so that $x_{k+j} \in B_\zeta(x^*)$ for $j \geq 0$. Then, conditions (3.1)–(3.4) and (3.7) give

$$\begin{aligned}
\|x_{k+1} - x^*\| &\leq \sum_{j=0}^{\infty} \|q_{k+j+1}\| \\
&\leq \Psi\Gamma \left(d(x_k, \mathcal{S})^2 + \sum_{j=1}^{\infty} \left(\Gamma d(x_k, \mathcal{S})\right)^{2^{j+1}-2} d(x_k, \mathcal{S})^2 \right) \\
&\leq \Psi\Gamma \left(1 + \sum_{j=1}^{\infty} (\Gamma\zeta)^{j-1} \right) d(x_k, \mathcal{S})^2 \\
&\leq 3\Psi\Gamma\, d(x_k, \mathcal{S})^2 \\
&\leq 3\Psi\Gamma\, \|x_k - x^*\|^2.
\end{aligned}$$

This shows the local Q-quadratic convergence of the sequence $\{x_k\}$.
□

4. Strong local convergence of ARQ and ARC. In this section we provide new results on the local convergence rate of ARQ and ARC methods to zero-residual solutions of (1.1). Under appropriate assumptions including the error bound condition on a limit point x^* in the set $\mathcal{S}$, we show that the sequence $\{x_k\}$ generated satisfies the conditions stated in Theorem 3.1. The results obtained are valid irrespective of the relation between the dimensions m and n of F.

In the following we make the following assumptions:

ASSUMPTION 4.1. *$F : \mathbb{R}^n \mapsto \mathbb{R}^m$ is continuously differentiable and for some solution $x^* \in \mathcal{S}$ there exists a constant $\epsilon > 0$ such that J is Lipschitz continuous with constant $2k_*$ in a neighbourhood $B_{2\epsilon}(x^*)$ of x^*.*

By Assumption 4.1 some technical results follow. The continuity of J implies that

$$\|J(x)\| \leq \kappa_J, \text{ for all } x \in B_{2\epsilon}(x^*) \text{ and some } \kappa_J > 0. \tag{4.1}$$

Moreover, for any point x, let $[x]_S \in \mathcal{S}$ be a vector such that $||x - [x]_s|| = d(x, \mathcal{S})$. Then, with $x^* \in \mathcal{S}$, we have that $[x_k]_S \in B_{2\epsilon}(x^*)$ whenever $x_k \in B_\epsilon(x^*)$, as $\|[x_k]_S - x^*\| \leq \|[x_k]_S - x_k\| + \|x_k - x^*\| \leq 2\|x_k - x^*\| \leq 2\epsilon$. As a consequence, by [11, Lemma 4.1.9]

$$\|F_k\| \leq \kappa_J \|\, [x_k]_S - x_k \,\|, \quad \text{if } x_k \in B_\epsilon(x^*), \tag{4.2}$$

and

$$\|F_k\| \leq \kappa_J \|\, x^* - x_k \,\|, \quad \text{if } x_k \in B_{2\epsilon}(x^*). \tag{4.3}$$

Further,

$$\|F(x+p) - F(x) - J(x)p\| \leq k_* \|p\|^2, \quad \text{for any } x \text{ and } x+p \text{ in } B_{2\epsilon}(x^*), \tag{4.4}$$

see [11, Lemma 4.1.12], and consequently

$$\|F_k + J_k([x_k]_S - x_k)\| \leq k_* d(x_k, \mathcal{S})^2, \quad \text{if } x_k \in B_\epsilon(x^*). \tag{4.5}$$

4.1. Analysis of the step. We prove that the step p_k generated by Algorithm 2.1 satisfies condition (3.1), i.e.

$$\|p_k\| \leq \Psi d(x_k, \mathcal{S}), \quad \text{if } \ x_k \in B_\epsilon(x^*), \tag{4.6}$$

for some strictly positive Ψ and ϵ.

We start showing that the minimizer p_k^* of the models satisfies

$$\|p_k^*\| \leq \Theta d(x_k, \mathcal{S}), \tag{4.7}$$

for some positive scalar Θ, if x_k is sufficiently close to x^*.

LEMMA 4.1. *Let Assumption 4.1 hold and $x^* \in \mathcal{S}$ be a limit point of the sequence $\{x_k\}$ generated by Algorithm 2.1. Suppose that $\sigma_k > \sigma_{\min} > 0$ for all $k \geq 0$. Then, if $x_k \in B_\epsilon(x^*)$, then p_k^* satisfies (4.7).*

Proof. Let $x_k \in B_\epsilon(x^*)$. Consider the ARQ method. Since p_k^* minimizes the model m_k^Q we have

$$m_k^Q(p_k^*) \leq m_k^Q([x_k]_S - x_k), \tag{4.8}$$

and by (4.5)

$$\begin{aligned} m_k^Q(p_k^*) &\leq \|F_k + J_k([x_k]_S - x_k)\| + \sigma_k \| [x_k]_S - x_k \|^2 \\ &\leq (k_* + \sigma_k) d(x_k, \mathcal{S})^2. \end{aligned} \tag{4.9}$$

Using (2.1) we get $\|p_k^*\|^2 \leq m_k^Q(p_k^*)/\sigma_k$, and the desired result follows from (4.9) and the assumption $\sigma_k > \sigma_{\min} > 0$.

For ARC method we proceed as above and get

$$\begin{aligned} m_k^C(p_k^*) &\leq \frac{1}{2}\|F_k + J_k([x_k]_S - x_k)\|^2 + \frac{1}{3}\sigma_k \|[x_k]_S - x_k\|^3 \\ &\leq \left(\frac{1}{2}k_*^2\epsilon + \frac{1}{3}\sigma_k\right) d(x_k, \mathcal{S})^3. \end{aligned}$$

Since (2.2) implies $\|p_k^*\|^3 \leq 3m_k^C(p_k^*)/\sigma_k$, the proof is completed. ∎

We now consider the case of practical interest where the minimizer of the model is approximately computed and characterize such an approximation. We proceed supposing that the couple (p_k, λ_k) satisfies the following two assumptions

ASSUMPTION 4.2. *The step p_k has the form $p_k = p(\lambda_k)$, i.e. it solves (2.3) with $\lambda = \lambda_k$.*

ASSUMPTION 4.3. *The scalar λ_k is such that*

$$\frac{\lambda_k^*}{1+\tau_k} \leq \lambda_k \leq \lambda_k^*(1+\tau_k),$$

for a given $\tau_k \in [0, \tau_{max}]$.

Trivially, these assumptions are satisfied when $p_k = p_k^*$. The upper bound τ_{max} on τ_k ensures that λ_k goes to zero as fast as λ_k^* does. In Section 4.3 we will show that, letting τ_k be a threshold chosen by the user, practical implementations of ARQ and ARC provide a step p_k satisfying both the above conditions.

The bound on the norm of p_k is now derived.

LEMMA 4.2. *Suppose that Assumption 4.1 holds and that* $x^* \in \mathcal{S}$ *is a limit point of the sequence* $\{x_k\}$ *generated by Algorithm 2.1. Let* $\sigma_k \geq \sigma_{\min} > 0$ *for all* $k \geq 0$. *Then there exists a positive* Ψ *such that if* $x_k \in B_\epsilon(x^*)$ *and* (p_k, λ_k) *satisfies Assumptions 4.2 and 4.3, then (4.6) holds.*

Proof. From Lemma 2.3 the function $\|p(\lambda)\|^2$ is monotonic decreasing for $\lambda \geq 0$. Then, when $\lambda_k \geq \lambda_k^*$ we have $\|p(\lambda_k)\| \leq \|p(\lambda_k^*)\|$ and by (4.7) the hypothesis follows. More generally, Assumption 4.3 yields

$$\|p(\lambda_k)\|^2 \leq \left\| p\left(\frac{\lambda_k^*}{1+\tau_k}\right)\right\|^2. \tag{4.10}$$

Moreover, from (2.7)

$$\begin{aligned}
\left\| p\left(\frac{\lambda_k^*}{1+\tau_k}\right)\right\|^2 &= \sum_{i=1}^{\ell} \frac{(\varsigma_i^k (r_k)_i)^2}{\left((\varsigma_i^k)^2 + \dfrac{\lambda_k^*}{1+\tau_k}\right)^2} \\
&= (1+\tau_k)^2 \sum_{i=1}^{\ell} \frac{(\varsigma_i^k (r_k)_i)^2}{\left((\varsigma_i^k)^2(1+\tau_k) + \lambda_k^*\right)^2} \\
&\leq (1+\tau_k)^2 \sum_{i=1}^{\ell} \frac{(\varsigma_i^k (r_k)_i)^2}{\left((\varsigma_i^k)^2 + \lambda_k^*\right)^2} \\
&= (1+\tau_k)^2 \|p(\lambda_k^*)\|^2.
\end{aligned}$$

Therefore, from (4.7) and $\tau_k \leq \tau_{max}$ we get the required result. ∎

4.2. Successful iterations and convergence of the sequence $\{x_k\}$. In order to apply Theorem 3.1, the second step is to prove that iteration k of ARQ and ARC is successful, i.e.

$$x_{k+1} = x_k + p_k, \quad \text{if } \; x_k \in B_\psi(x^*), \tag{4.11}$$

for some $\psi \in (0, \epsilon)$.

In the rest of the section the error bound Assumption 2.1 is supposed on the limit point x^* and the scalar ϵ in Assumption 4.1 is possibly reduced to be such that

$$\epsilon \leq \chi,$$

where χ is as in (2.14). Moreover we require that

$$\sigma_k \leq \sigma_{\max} \quad \text{for all} \quad k \geq 0, \tag{4.12}$$

for some positive $\sigma_{\max}$.

In the results below we will make use of the interpretation of p_k as the minimizer of the model m_k^{LM} given in (2.13) with $\mu_k = \lambda_k$. Thus, by (4.5)

$$\begin{aligned} m_k^{LM}(p_k) &\leq m_k^{LM}([x_k]_S - x_k) \\ &= \frac{1}{2}\|F_k + J_k([x_k]_S - x_k)\|^2 + \frac{1}{2}\lambda_k\|[x_k]_S - x_k\|^2 \\ &\leq \frac{1}{2}k_*^2\|[x_k]_S - x_k\|^4 + \frac{1}{2}\lambda_k\|[x_k]_S - x_k\|^2, \end{aligned} \tag{4.13}$$

whenever $x_k \in B_\epsilon(x^*)$.

LEMMA 4.3. *Let Assumptions 4.1, 4.2 and 4.3 hold and $x^* \in \mathcal{S}$ be a limit point of the sequence $\{x_k\}$ generated by the ARQ method satisfying Assumption 2.1. Suppose that $\sigma_{\max} \geq \sigma_k \geq \sigma_{\min} > 0$ for all $k \geq 0$. Then there exist positive scalars ψ and Λ such that, if $x_k \in B_\psi(x^*)$,*

$$m_k^Q(p_k) \leq \Lambda d(x_k, S)^{3/2}, \tag{4.14}$$

and iteration k is very successful.

Proof. Let $\psi = \epsilon/(1+\Psi)$ where ϵ and Ψ are the scalars in Lemma 4.2. Assume that $x_k \in B_\psi(x^*)$. Using (4.13), Assumption 4.3, (2.5) and (4.2)

$$m_k^{LM}(p_k) \leq \left(\frac{1}{2}k_*^2\psi + \sigma_k\kappa_{\text{J}}(1+\tau_{max})\right) d(x_k, \mathcal{S})^3.$$

Then, using (2.13)

$$\|J_k p_k + F_k\| \leq \sqrt{2m_k^{LM}(p_k)} \leq \sqrt{k_*^2\psi + 2\sigma_k\kappa_{\text{J}}(1+\tau_{max})}\; d(x_k, \mathcal{S})^{3/2},$$

and by (2.1) and (4.6)

$$m_k^Q(p_k) \leq \sqrt{k_*^2\psi + 2\sigma_k\kappa_{\text{J}}(1+\tau_{max})}\, d(x_k, \mathcal{S})^{3/2} + \sigma_k\Psi^2 d(x_k, \mathcal{S})^2.$$

This latter inequality and (4.12) yield (4.14).

Finally, we show that iteration k is very successful. Since (4.6) gives

$$\|x_k + p_k - x^*\| \leq \|x_k - x^*\| + \|p_k\| \leq \psi + \Psi d(x_k, \mathcal{S}) \leq \psi(1+\Psi), \tag{4.15}$$

from the definition of ψ, it follows that $x_k + p_k \in B_\epsilon(x^*)$. Using (4.4) the quantity $\|F(x_k + p_k)\|$ can be bounded as

$$\|F(x_k + p_k)\| \leq \|F(x_k + p_k) - F_k - J_k p_k\| + \|F_k + J_k p_k\| \tag{4.16}$$

$$\leq k_*\|p_k\|^2 + m_k^Q(p_k) \tag{4.17}$$

Thus, condition (2.10) can be bounded below as

$$\rho_k = 1 - \frac{\|F(x_k + p_k)\| - m_k^Q(p_k)}{\|F_k\| - m_k^Q(p_k)} \geq 1 - \frac{\kappa_*\|p_k\|^2}{\|F_k\| - m_k^Q(p_k)}.$$

Moreover, (4.14) and (2.14) yield

$$\|F_k\| - m_k^Q(p_k) \geq \|F_k\| - \Lambda d(x_k, \mathcal{S})^{3/2} \geq (1 - \alpha^{3/2}\Lambda\|F_k\|^{1/2})\|F_k\|,$$

and the last expression is positive if $\|F_k\|$ is small enough, i.e. if x_k is sufficiently close to x^*. Thus, by (4.6) and (2.14) we obtain

$$\rho_k \geq 1 - \frac{k_* \alpha^2 \Psi^2 \|F_k\|}{1 - \alpha^{3/2} \Lambda \|F_k\|^{1/2}},$$

and reducing ψ, if necessary, we get $\rho_k \geq \eta_2$ for the fixed value $1 > \eta_2 > 0$.
□

Regarding the ARC algorithm, we have an analogous result shown in the next lemma.

LEMMA 4.4. *Let Assumptions 4.1, 4.2 and 4.3 hold and $x^* \in \mathcal{S}$ is a limit point of the sequence $\{x_k\}$ generated by the ARC method satisfying Assumption 2.1. Suppose that $\sigma_{\max} \geq \sigma_k \geq \sigma_{\min} > 0$ for all $k \geq 0$. Then there exist positive scalars ψ and Λ such that, if $x_k \in B_\psi(x^*)$,*

$$m_k^C(p_k) \leq \Lambda d(x_k, \mathcal{S})^3, \tag{4.18}$$

and iteration k is very successful.

Proof. Let $\psi = \epsilon/(1+\Psi)$ where ϵ and Ψ are the scalars in Lemma 4.2. Assume that $x_k \in B_\psi(x^*)$. By (4.13), (2.6), (4.7) and Assumption 4.3

$$m_k^{LM}(p_k) \leq \frac{1}{2}\left(k_*^2 \psi + \sigma_k \Theta(1 + \tau_{max})\right) d(x_k, \mathcal{S})^3.$$

Consequently, by the definition of m_k^C, inequality $\frac{1}{2}\|J_k p_k + F_k\|^2 \leq m_k^{LM}(p_k)$ and (4.6) we get

$$m_k^C(p_k) \leq \frac{1}{2}\left(k_*^2 \psi + \sigma_k \Theta(1 + \tau_{max})\right) d(x_k, \mathcal{S})^3 + \frac{1}{3}\sigma_k \Psi^3 d(x_k, \mathcal{S})^3$$

and (4.18) follows from (4.12).

Let now focus on very successful iterations. Proceeding as in Lemma 4.3 we can derive (4.15), i.e. $x_k + p_k \in B_\epsilon(x^*)$. Thus, by using (4.6), (4.4) and (4.18) we have

$$\begin{aligned}
\|F(x_k + p_k)\|^2 &\leq \|F(x_k + p_k) - F_k - J_k p_k\|^2 + \|F_k + J_k p_k\|^2 + 2(F_k + J_k p_k)^T (F(x_k + p_k) - F_k - J_k p_k) \\
&\leq k_*^2 \|p_k\|^4 + 2m_k^C(p_k) + 2\|F(x_k + p_k) - F_k - J_k p_k\| \|F_k + J_k p_k\| \\
&\leq k_*^2 \Psi^4 d(x_k, \mathcal{S})^4 + 2m_k^C(p_k) + 2k_* \Psi^2 \sqrt{2\Lambda} d(x_k, \mathcal{S})^{7/2},
\end{aligned}$$

i.e., there exists a constant Φ such that

$$\frac{1}{2}\|F(x_k + p_k)\|^2 - m_k^C(p_k) \leq \Phi d(x_k, \mathcal{S})^{7/2}.$$

Consequently, ρ_k in (2.11) can be bounded below as

$$\rho_k = 1 - \frac{\frac{1}{2}\|F(x_k + p_k)\|^2 - m_k^C(p_k)}{\frac{1}{2}\|F_k\|^2 - m_k^C(p_k)} \geq 1 - \frac{\Phi d(x_k, \mathcal{S})^{7/2}}{\frac{1}{2}\|F_k\|^2 - m_k^C(p_k)},$$

and (4.18) and (2.14) give

$$\frac{1}{2}\|F_k\|^2 - m_k^C(p_k) \geq \frac{1}{2}\|F_k\|^2 - \Lambda d(x_k, \mathcal{S})^3 \geq \|F_k\|^2 \left(\frac{1}{2} - \alpha^3 \Lambda \|F_k\|\right),$$

where the last expression is positive if x_k is close enough to x^*. Thus, by (2.14)

$$\rho_k \geq 1 - \frac{\alpha^{7/2}\Phi\|F_k\|^{3/2}}{\frac{1}{2} - \alpha^3\Lambda\|F_k\|},$$

and reducing ψ, if necessary, we get $\rho_k \geq \eta_2$ for the fixed value $1 > \eta_2 > 0$. ∎

The next lemma establishes the dependence of $d(x_k + p_k, \mathcal{S})$ upon $d(x_k, \mathcal{S})$ whenever x_k is sufficiently close to x^*.

LEMMA 4.5. *Let Assumptions 2.1, 4.1 and 4.2 hold. Suppose that p_k satisfies (4.6) and*

$$\lambda_k \leq d(x_k, \mathcal{S})^{\xi},$$

with $\xi \in (0, 2]$. Then, if x_k is close enough to x^*

$$d(x_k + p_k, \mathcal{S}) \leq \Gamma d(x_k, \mathcal{S})^{\min\{\xi+1,\, 2\}}, \tag{4.19}$$

for some positive Γ.

Proof. The proof is given in [1, Lemma 4]. ∎

With the previous results at hand, exploiting Theorem 3.1, we are ready to show the local convergence behaviour of both adaptive regularized approaches.

COROLLARY 4.6. *Let Assumptions 4.1, 4.2 and 4.3 hold and $x^* \in \mathcal{S}$ be a limit point of the sequence $\{x_k\}$ generated by the ARQ method satisfying Assumption 2.1. Suppose that $\sigma_{\max} \geq \sigma_k \geq \sigma_{\min} > 0$ for all $k \geq 0$. Then, $\{x_k\}$ converges to x^* Q-quadratically.*

Proof. Lemmas 4.2 and 4.3 guarantee conditions (3.1) and (3.2). Moreover, by Assumption 4.3, (2.5) and (4.2)

$$\lambda_k \leq \lambda_k^*(1 + \tau_{max}) \leq 2(1 + \tau_{max})\sigma_k\|F_k\| \leq 2(1 + \tau_{max})\sigma_{max}\kappa_J\, d(x_k, \mathcal{S}).$$

Therefore, by Lemma 4.5 we have that (4.19) holds with $\xi = 1$ and such an inequality coincides with (3.3). Then, the proof is completed by using Theorem 3.1. ∎

COROLLARY 4.7. *Let Assumptions 4.1, 4.2 and 4.3 hold and $x^* \in \mathcal{S}$ be a limit point of the sequence $\{x_k\}$ generated by the ARC method satisfying Assumption 2.1. Suppose that $\sigma_{\max} \geq \sigma_k \geq \sigma_{\min} > 0$ for all $k \geq 0$. Then, $\{x_k\}$ converges to x^* Q-quadratically.*

Proof. Lemmas 4.2 and 4.4 guarantee conditions (3.1) and (3.2). Further, by Assumption 4.3, (2.6) and (4.7)

$$\lambda_k \leq \lambda_k^*(1 + \tau_{max}) \leq (1 + \tau_{max})\sigma_k\, \|p_k^*\| \leq (1 + \tau_{max})\sigma_{max}\Theta\, d(x_k, \mathcal{S}).$$

Therefore, by Lemma 4.5 the inequality (4.19) holds with $\xi = 1$, i.e. (3.3) is met. Then, the proof is completed by using Theorem 3.1. ∎

The previous results have been obtained supposing that $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$. The lower bound $\sigma_k \geq \sigma_{\min}$ can be straightforwardly enforced in the algorithm for some small specified threshold $\sigma_{\min}$. Concerning condition $\sigma_k \leq \sigma_{\max}$, we now discuss when it is satisfied. Focusing on the ARQ method, in [2, Lemma 4.7] it has been proved that (4.12) holds under the following two assumptions on $J(x)$: *i)* $\|J(x)\|$ is uniformly

bounded above for all $k \geq 0$ and all $x \in [x_k, x_k + p_k]$; *ii*) there exist positive constants $\kappa_{\rm L}, \kappa_{\rm S}$ such that, if $\|x - x_k\| \leq \kappa_{\rm S}$ and $x \in [x_k, x_k + p_k]$, then $\|J(x) - J_k\| \leq \kappa_{\rm L}\|x - x_k\|$ for all $k \geq 0$.

For the ARC method, suppose that F is twice continuously differentiable and the Hessian matrix H of f is globally Lipschitz continuous in $\mathbb{R}^n$, with Lipschitz constant $\kappa_{\rm H}$. We now show two occurrences where (4.12) holds. By [4, Lemma 5.2], (4.12) is guaranteed if

$$\|(H_k - J_k^T J_k)p_k\| \leq C\|p_k\|^2, \tag{4.20}$$

for all $k \geq 0$ and some constant $C > 0$. Since $H_k - J_k^T J_k = \sum_{i=1}^m F_i(x_k)\nabla^2 F_i(x_k)$, where F_i is the i-th component of F, $1 \leq i \leq m$, then (4.20) is satisfied provided that $\|F_k\| \leq \kappa_{\rm F}\|p_k\|$ for some $\kappa_{\rm F} > 0$ and all $k \geq 0$. Alternatively, (4.12) holds if $x_k \to x^*$. In fact,

$$f(x_k + p_k) - m_k(p_k) = \frac{1}{2} p_k^T \left(H(\zeta_k) - J_k^T J_k\right) p_k - \frac{\sigma_k}{3}\|p_k\|^3,$$

for some ζ_k on the line segment $(x_k, x_k + p_k)$, see [4, Equation (4.2)]. Then,

$$\begin{aligned} f(x_k + p_k) - m_k(p_k) &\leq \frac{1}{2}\|H(\zeta_k) - H(x_k)\|\,\|p_k\|^2 + \frac{1}{2}\|(H(x_k) - J_k^T J_k)p_k\|\,\|p_k\| - \frac{\sigma_k}{3}\|p_k\|^3 \\ &\leq \frac{1}{2}\kappa_{\rm H}\|p_k\|^3 + \tilde{\kappa}_{\rm H}\|F_k\|\,\|p_k\|^2 - \frac{\sigma_k}{3}\|p_k\|^3, \end{aligned}$$

for some positive $\tilde{\kappa}_{\rm H}$. Thus, for all k sufficiently large (4.6) and (2.14) yield

$$f(x_k + p_k) - m_k(p_k) \leq \left(\alpha\Psi\kappa_{\rm H} + \tilde{\kappa}_{\rm H} - \alpha\Psi\frac{\sigma_k}{3}\right)\alpha^2\Psi^2\|F_k\|^3,$$

and for $\sigma_k > 3(\alpha\Psi\kappa_{\rm H} + \tilde{\kappa}_{\rm H})/(\alpha\Psi)$ the iteration is very successful. Consequently, the updating rule (2.12) gives $\sigma_{k+1} \leq \sigma_k$ and (4.12) follows.

Summarizing, we have shown convergence results for ARQ and ARC that are analogous to results known in literature for the Levenberg-Marquardt methods when the parameter μ_k has the form (2.15). Concerning the choice of σ_k and μ_k, we underline that the rule for fixing σ_k is simpler to implement than the rule for choosing μ_k. In fact, σ_k is fixed on the base of the adaptive choice in Algorithm 2.1 while (2.15) leaves the choice of both δ and the constant multiplying $\|F_k\|$ open and this may have an impact on the practical behaviour of the Levenberg-Marquardt methods [17, 18].

Finally, in [5] the model $m_k^C(p)$ has been generalized to

$$m_k^{2,\beta}(p) = \frac{1}{2}\|F_k + J_k p\|^2 + \frac{1}{\beta}\sigma_k\|p\|^\beta, \tag{4.21}$$

with $\beta \in [2, 3]$. Trivially, $m_k^{2,\beta}(p)$ reduces to $m_k^C(p)$ when $\beta = 3$. The adaptive regularized procedure based on the use of $m_k^{2,\beta}(p)$ can be analyzed using the same arguments as above. Assume the same assumptions as in Corollary 4.7. Then the adaptive procedure converges to $x^* \in \mathcal{S}$ superlinearly when $\beta \in (2, 3)$. In fact, (4.6) holds and a slight modification of Lemma 4.4 shows that eventually all iterations are very successful. By [5], $\lambda_k^* = \sigma_k \|p_k^*\|^{\beta-2}$ and proceeding as in Corollary 4.7

$$\lambda_k \leq (1 + \tau_{max})\sigma_{max}\,\Theta^{\beta-2} d(x_k, \mathcal{S})^{\beta-2}.$$

Thus, Lemma 4.5 implies $d(x_k + p_k, \mathcal{S}) \leq \Gamma d(x_k, \mathcal{S})^{\beta-1}$, and a straightforward adaptation of Theorem 3.1 yields Q-superlinear convergence with rate $\beta - 1$.

4.3. Computing the trial step. In this section we consider a viable way devised in [2, 4] for computing an approximate minimizer of m_k and enforcing Assumptions 4.2 and 4.3. In such an approach a couple (p_k, λ_k) satisfying

$$p_k = p(\lambda_k), \qquad (J_k^T J_k + \lambda_k I)p_k = -g_k, \tag{4.22}$$

with λ_k being an approximation to λ_k^*, is sought. The scalar λ_k can be obtained applying a root-finding solver to the so-called secular equation. In fact, from Lemma 2.2, the optimal scalar λ_k^* for ARQ solves the scalar nonlinear equation

$$\rho(\lambda) = \lambda - 2\sigma_k \|J_k p(\lambda) + F_k\| = 0. \tag{4.23}$$

The function $\rho'(\lambda)$ may change sign in $(0, +\infty)$, while the reformulation

$$\psi^Q(\lambda) = -\frac{\rho(\lambda)}{\lambda} = 0, \tag{4.24}$$

is such that $\psi^Q(\lambda)$ is convex and strictly decreasing in $(0, +\infty)$ [2]. Analogously, in ARC the scalar λ^* solves the scalar nonlinear equation

$$\rho(\lambda) = \lambda - \sigma_k \|p(\lambda)\| = 0, \tag{4.25}$$

which can be reformulated as

$$\psi^C(\lambda) = -\frac{\rho(\lambda)}{\lambda \|p(\lambda)\|} = 0. \tag{4.26}$$

The function $\psi^C(\lambda)$ is convex and strictly decreasing in $(0, +\infty)$ [4]. In what follows we will use the notation $\psi(\lambda)$ in all the expressions that holds for both functions.

Due to the monotonicity and convexity properties of $\psi(\lambda)$, either the Newton or the secant method applied to (4.24) and (4.26) converges globally and monotonically to the positive root λ_k^* for any initial guess in $(0, \lambda_k^*)$. We refer to [2] and [4] for details on the evaluation of $\psi(\lambda)$ and its first derivatives.

Clearly p_k satisfies Assumption 4.2, while Assumption 4.3 is met if a suitable stopping criterion is imposed to the root-finding solver. Let the initial guess be $\lambda_k^0 \in (0, \lambda_k^*)$. Then, the sequence $\{\lambda_k^\ell\}$ generated is such that $\lambda_k^\ell > \lambda_k^0$ for any $\ell > 0$ and by Taylor expansion there exists $\bar{\lambda} \in (\lambda_k^\ell, \lambda_k^*)$ such that

$$\psi(\lambda_k^\ell) = \psi'(\bar{\lambda})(\lambda_k^\ell - \lambda_k^*), \quad \text{i.e.} \quad \lambda_k^* - \lambda_k^\ell = -\frac{\psi(\lambda_k^\ell)}{\psi'(\bar{\lambda})}.$$

In principle, if the iterative process is stopped when

$$\psi(\lambda_k^\ell) < -\tau_k \lambda_k^\ell \psi'(\bar{\lambda}), \tag{4.27}$$

then the couple $(p(\lambda_k^\ell), \lambda_k^\ell)$ satisfies Assumptions 4.2, 4.3. A practical implementation of this stopping criterion can be carried out by using an upper bound $\bar{\lambda}_U$ on $\bar{\lambda}$. Since $\psi(\lambda)$ is convex it follows that $\psi'(\lambda)$ is strictly increasing, $\psi'(\bar{\lambda}) \leq \psi'(\bar{\lambda}_U)$ and (4.27) is guaranteed by enforcing

$$\psi(\lambda_k^\ell) < -\tau_k \lambda_k^\ell \psi'(\bar{\lambda}_U).$$

Possible choices for $\bar{\lambda}_U$ follows from using the condition $\bar{\lambda} \leq \lambda_k^*$. In particular, in ARQ method Lemma 2.2 suggests $\bar{\lambda}_U = 2\sigma_k \|F_k\|$. In ARC method, Lemma 2.2 and the monotonic decrease of $\|p(\lambda)\|$ in $(0, \infty)$, shown in Lemma 2.3, yield $\bar{\lambda}_U = \sigma_k \|p(\lambda_k^\ell)\|$. Finally we note that if the bisection process is used to get an initial guess for the Newton process, then $\bar{\lambda}_U$ can be taken as the right extreme of the last bracketing interval computed.

5. Computing the trial step in a subspace. The strategy for computing the trial step devised in the previous section requires the solution of a sequence of linear systems of the form (4.22). Namely, for each value of λ_k^ℓ generated by the root-finding solver applied to the secular equation, the computation of $\psi(\lambda_k^\ell)$ is needed and this calls for the solution of (4.22).

A different approach can be used when large scale problems are solved and the factorization of coefficient matrix of (4.22) is unavailable due to cost or memory limitations. In such an approach the model m_k is minimized over a sequence of nested subspaces. The Golub and Kahan bi-diagonalization process [5] is used to generate such subspaces and minimizing the model in the subspaces is quite inexpensive. The minimization process is carried out until a step p_k^s satisfying

$$\|\nabla_p m_k(p_k^s)\| \leq \omega_k, \tag{5.1}$$

is computed for some positive tolerance ω_k. We now study the effect of using the step $p_k = p_k^s$ in Algorithm 2.1.

At termination of the iterative process the couple (λ_k, p_k^s) satisfies:

$$(J_k^T J_k + \lambda_k I)p_k^s + g_k = r_k \tag{5.2}$$

$$\lambda_k - \phi(p_k^s) = 0 \tag{5.3}$$

where

$$\phi(p_k^s) = \begin{cases} 2\sigma_k\|F_k + J_k p_k^s\| & \text{when } m_k = m_k^Q \\ \sigma_k\|p_k^s\| & \text{when } m_k = m_k^C \end{cases} .$$

¿From condition (5.1) and the form of $\nabla_p m_k$ it follows that

$$\|r_k\| \leq \bar{\omega}_k, \tag{5.4}$$

where

$$\bar{\omega}_k = \begin{cases} \omega_k\|F_k + J_k p_k^s\| & \text{when } m_k = m_k^Q \\ \omega_k & \text{when } m_k = m_k^C \end{cases} . \tag{5.5}$$

Since p_k^s does not satisfy Assumption 4.2 some properties of the approximate minimizer p_k discussed in Section 4.3 are not shared by p_k^s and we cannot rely on the convergence theory developed in the previous section.

The relation between p_k^s and $p(\lambda_k)$ follows from noting that (5.2) yields

$$(J_k^T J_k + \lambda_k I)p_k^s = (J_k^T J_k + \lambda_k I)p(\lambda_k) + r_k,$$

i.e.,

$$p_k^s - p(\lambda_k) = (J_k^T J_k + \lambda_k I)^{-1} r_k. \tag{5.6}$$

On the other hand, Assumption 4.3 can be satisfied choosing the tolerance ω_k in (5.1) small enough.

In this section we show when the strong local convergence behaviour of the ARQ and ARC procedures is retained if p_k^s is used. We start studying the quadratic model and the step generated by the ARQ method.

LEMMA 5.1. *Let $\{x_k\}$ be the sequence generated by the ARQ method with steps $p_k = p_k^s$ satisfying (5.1)–(5.3). Suppose that Assumptions 4.1 and 4.3 hold and $x^* \in \mathcal{S}$ is a limit point of $\{x_k\}$. If $\sigma_k \geq \sigma_{\min} > 0$ for all $k \geq 0$, and*

$$\omega_k \leq \theta \|F_k\|^{3/2}, \tag{5.7}$$

for a positive θ, then there exists a positive $\bar{\Psi}$ such that,

$$\|p_k^s\| \leq \bar{\Psi} d(x_k, \mathcal{S}), \tag{5.8}$$

whenever x_k is sufficiently close to x^*. Moreover, if (4.12) holds then there exists a positive $\bar{\Lambda} > 0$ such that

$$m_k^Q(p_k^s) \leq \bar{\Lambda} d(x_k, \mathcal{S})^{3/2}, \tag{5.9}$$

whenever x_k is sufficiently close to x^*.

Proof. Let $x_k \in B_\epsilon(x^*)$. By (5.4), (5.5) and (5.3) we get

$$\|(J_k^T J_k + \lambda_k I)^{-1} r_k\| \leq \frac{1}{\lambda_k} \bar{\omega}_k \leq \frac{1}{2\sigma_{\min}} \omega_k,$$

and using (5.7) and (4.2)

$$\|(J_k^T J_k + \lambda_k I)^{-1} r_k\| \leq \frac{\theta \kappa_J^{3/2}}{2\sigma_{\min}} d(x_k, \mathcal{S})^{3/2}. \tag{5.10}$$

Since $\|p(\lambda_k)\|$ satisfies (4.6) by Lemma 4.2, and (5.6) gives $\|p_k^s\| \leq \|p(\lambda_k)\| + \|(J_k^T J_k + \lambda_k I)^{-1} r_k\|$, inequality (5.8) holds.

Let us now focus on inequality (5.9). Assuming $x_k \in B_\psi(x^*)$ where ψ is the scalar in Lemma 4.3, and using (4.1), (4.14) and (4.12) we obtain

$$\begin{aligned} m_k^Q(p_k^s) &\leq \|F_k + J_k p(\lambda_k)\| + \|J_k(p_k^s - p(\lambda_k))\| + \sigma_k \|p_k^s\|^2 \\ &\leq \Lambda d(x_k, \mathcal{S})^{3/2} + \kappa_J \|p_k^s - p(\lambda_k)\| + \sigma_{max} \bar{\Psi}^2 d(x_k, \mathcal{S})^2, \end{aligned}$$

and by (5.6) and (5.10)

$$m_k^Q(p_k^s) \leq \Lambda d(x_k, \mathcal{S})^{3/2} + \frac{\theta \kappa_J^{5/2}}{2\sigma_{\min}} d(x_k, \mathcal{S})^{3/2} + \sigma_{max} \bar{\Psi}^2 d(x_k, \mathcal{S})^2,$$

which completes the proof. □

The following Lemma shows the corresponding result for the ARC method.

LEMMA 5.2. *Let $\{x_k\}$ be the sequence generated by the ARC method with steps $p_k = p_k^s$ satisfying (5.1)–(5.3). Suppose that Assumptions 4.1 and 4.3 hold and $x^* \in \mathcal{S}$ is a limit point of the sequence $\{x_k\}$. If $\sigma_k \geq \sigma_{\min} > 0$ for all $k \geq 0$, and*

$$\omega_k \leq \theta_k \|F_k\|^{3/2}, \quad \theta_k = \kappa_\theta \min(1, \|p_k^s\|), \tag{5.11}$$

for a positive scalar κ_θ, then there exists a positive $\bar{\Psi}$ such that (5.8) holds for x_k sufficiently close to x^*. Moreover, if (4.12) holds then there exists $\bar{\Lambda} > 0$ such that

$$m_k^C(p_k^s) \leq \bar{\Lambda} d(x_k, \mathcal{S})^3, \tag{5.12}$$

whenever x_k is sufficiently close to x^*.

Proof. Let $x_k \in B_\epsilon(x^*)$. First, note that by (4.2), (5.3), (5.4), (5.5) and (5.11)

$$\begin{aligned} \|(J_k^T J_k + \lambda_k I)^{-1} r_k\| &\leq \frac{\bar{\omega}_k}{\lambda_k} \leq \frac{1}{\sigma_k \|p_k^s\|} \bar{\omega}_k, \\ &\leq \frac{\kappa_\theta \kappa_{\mathrm{J}}^{3/2}}{\sigma_{\min}} d(x_k, \mathcal{S})^{3/2}. \end{aligned} \tag{5.13}$$

Since $\|p(\lambda_k)\|$ satisfies (4.6) by Lemma 4.2, and (5.6) gives $\|p_k^s\| \leq \|p(\lambda_k)\| + \|(J_k^T J_k + \lambda_k I)^{-1} r_k\|$, (5.8) holds.

Moreover, letting $x_k \in S_\psi(x^*)$, and using (4.1), (4.18), (5.6) and (5.13) we get

$$\begin{aligned} m_k^C(p_k^s) &\leq \frac{1}{2}\|F_k + J_k p(\lambda_k)\|^2 + \frac{1}{2}\|J_k(p_k^s - p(\lambda_k))\|^2 \\ &\quad + (F_k + J_k p(\lambda_k))^T J_k (p_k^s - p(\lambda_k)) + \frac{1}{3}\sigma_k \|p_k^s\|^3 \\ &\leq \Lambda d(x_k, \mathcal{S})^3 + \frac{\kappa_\theta^2 \kappa_{\mathrm{J}}^5}{2\sigma_{\min}^2} d(x_k, \mathcal{S})^3 + \frac{\kappa_\theta \kappa_{\mathrm{J}}^{5/2} \sqrt{2\Lambda}}{\sigma_{\min}} d(x_k, \mathcal{S})^3 + \frac{1}{3}\sigma_{\max} \bar{\Psi}^3 d(x_k, \mathcal{S})^3, \end{aligned}$$

and this yields (5.12) □

We are now able to state the local convergence results for ARQ and ARC methods in the case where the step is computed in a subspace.

COROLLARY 5.3. *Let $\{x_k\}$ be the sequence generated by the ARQ method with steps $p_k = p_k^s$ satisfying (5.1)–(5.3). Suppose Assumptions 4.1 and 4.3 hold and that $x^* \in \mathcal{S}$ is a limit point of $\{x_k\}$ satisfying Assumption 2.1. If $0 < \sigma_{\min} \leq \sigma_k \leq \sigma_{\max}$ for all $k \geq 0$ and (5.7) holds, then $\{x_k\}$ converges to x^* Q-quadratically.*

Proof. Using (5.8) and (5.9) and proceeding as in the proof of Lemma 4.3, we can prove that iteration k is very successful whenever x_k is sufficiently close to x^*. Moreover, as (5.1) and (5.7) are satisfied, Lemma 4 in [1] guarantees that inequality (3.3) holds and Theorem 3.1 yields the hypothesis. □

COROLLARY 5.4. *Let $\{x_k\}$ be the sequence generated by the ARC method with steps $p_k = p_k^s$ satisfying (5.1)–(5.3). Assume that Assumptions 4.1 and 4.3 hold and $x^* \in \mathcal{S}$ is a limit point of $\{x_k\}$ satisfying Assumption 2.1. If $0 < \sigma_{\min} \leq \sigma_k \leq \sigma_{\max}$ for all $k \geq 0$ and (5.11) holds, then $\{x_k\}$ converges to x^* Q-quadratically.*

Proof. Using (5.8) and (5.12) and proceeding as in the proof of Lemma 4.4, we can prove that iteration k is very successful whenever x_k is sufficiently close to x^*. Moreover, as (5.1) and (5.11) are satisfied, Lemma 4 in [1] guarantees that inequality (3.3) holds and Theorem 3.1 yields the hypothesis. □

6. Conclusion. In this paper, we have studied the local convergence behaviour of two adaptive regularized methods for solving nonlinear least-squares problems and we have established local quadratic convergence to zero-residual solutions under an error bound assumption. Interestingly, this condition is considerably weaker than the standard assumptions used in literature and the results obtained are valid for under and over-determined problems as well as for square problems.

The theoretical analysis carried out shows that the regularizations enhance the properties of the underlying unregularized methods. The focus on zero-residual solutions is a straightforward consequence of models considered which are regularized Gauss-Newton models. Further results on potentially rank-deficient nonliner least-squares have been given in [8] for adaptive cubic regularized models employing suitable approximations of the Hessian of the objective function.

REFERENCES

[1] R. Behling and A. Fischer. A unified local convergence analysis of inexact constrained Levenberg-Marquardt methods. *Optimization Letters*, **6**, 927–940, 2012.

[2] S. Bellavia, C. Cartis, N. I. M. Gould, B. Morini, and Ph. L. Toint. Convergence of a Regularized Euclidean Residual Algorithm for Nonlinear Least-Squares. *SIAM Journal on Numerical Analysis*, **48**, 1–29, 2010.

[3] H. Y. Benson and D. F. Shanno. Interior-Point methods for nonconvex nonlinear programming: cubic regularization. Technical report, 2012, http://www.optimization-online.org/DB_FILE/2012/02/3362.pdf.

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

[5] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Trust-region and other regularisations of linear least-squares problems. *BIT Numerical Mathematics*, **49**, 21-53, 2009

[6] C. Cartis, N. Gould and Ph. L. Toint. Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity. *Mathematical Programming A*, **30**, 295-319, 2011.

[7] C. Cartis, N. Gould and Ph. L. Toint. An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity. *IMA Journal on Numerical Analysis*, **4**, 1662–1695, 2012.

[8] C. Cartis, N. I. M. Gould, and Ph. L. Toint. On the evaluation complexity of cubic regularization methods for potentially rank-deficient nonlinear least-squares problems and its relevance to constrained nonlinear optimization. SIAM Journal on Optimization, **23**, pp. 1553–1574, 2013.

[9] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. No. 1 in the 'MPS–SIAM series on optimization'. SIAM, Philadelphia, USA, 2000.

[10] H. Dan, N. Yamashita and M. Fukushima. Convergence properties of the inexact Levenberg-Marquardt method under local error bound conditions. *Optimization Methods and Software*, **17**, 605-626, 2002.

[11] J.E. Dennis, R.B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations Prentice Hall, Englewood Cliffs, NJ, 1983.

[12] J. Fan and J. Pan. Inexact Levenberg-Marquardt method for nonlinear equations. *Discrete and Continuous Dynamical Systems, Series B*, **4**, 1223–232, 2004.

[13] J. Fan, Y. Yuan. On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption *Computing*, **74**, 23–39, 2005.

[14] N. I. M. Gould, M. Porcelli, and Ph. L. Toint. Updating the regularization parameter in the adaptive cubic regularization algorithm. *Computational Optimization and Applications*, **53**, 1–22, 2012.

[15] A. Griewank. The modification of Newton's method for unconstrained optimization by bounding cubic terms. Technical Report NA/12 (1981), Department of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom, 1981.

[16] W. W. Hager and H. Zhang, Self-adaptive inexact proximal point methods. *Computational Optimization and Applications*, **39**, 161–181, 2008.

[17] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints, *Journal of Computational and Applied Mathematics*, **172**, 375–397, 2004.

[18] M. Macconi, B. Morini and M. Porcelli, Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities. *Applied Numerical Mathematics*, **59**, 859–876, 2009.

[19] J.J. Moré. The Levenberg-Marquardt algorithm: implementation and theory. *Lecture Notes in Mathematics*, **630**, 105-116, Springer, Berlin, 1978.

[20] Yu. Nesterov. Modified Gauss-Newton scheme with worst-case guarantees for global performance. *Optimization Methods and Software*, **22**, 469–483, 2007.

[21] Yu. Nesterov and B. T. Polyak. Cubic regularization of Newton's method and its global performance. *Mathematical Programming*, **108**, 177–205, 2006.

[22] J.S. Pang, Error bounds in mathematical programming *Mathematical Programming*, **79**, 299–332, 1997.

[23] Ph. L. Toint. Nonlinear Stepsize Control, Trust Regions and Regularizations for Unconstrained Optimization. *Optimization Methods and Software*, **28**, 82–95, 2013.

[24] P. Tseng. Error bounds and superlinear convergence analysis of some Newton-type methods in optimization. in *Nonlinear Optimization and Applications, 2*, eds. G. Di Pillo and F. Giannessi, Kluwer Academic Publishers, Dordrecht Appl. Optim. 36 , 445–462, 2000.

[25] M. Weiser, P. Deuflhard, and B. Erdmann. Affine conjugate adaptive Newton methods for nonlinear elastomechanics. *Optimization Methods and Software*, **22**, 413–431, 2007.

[26] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg-Marquardt method. *Computing Supplementa*, **15**, 237–249, 2001.

[27] H. Zhang and A. R. Conn. On the local convergence of a derivative-free algorithm for least-squares minimization. *Computational Optimization and Applications*, **51**, 481–507,2012.

[28] D. Zhu. Affine scaling interior Levenberg-Marquardt method for bound-constrained semismooth equations under local error bound conditions. *Journal of Computational and Applied Mathematics*, **219**, 198–216, 2008.