To appear in *Applicable Analysis*
Vol. 00, No. 00, Month 20XX, 1–29

On the second order asymptotical regularization of linear ill-posed inverse problems

Y. Zhang[a,b*], B. Hofmann[a]

[a] *Faculty of Mathematics, Chemnitz University of Technology, Chemnitz 09107, Germany*;
[b] *School of Science and Technology, Örebro University, Örebro 70182, Sweden*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

In this paper, we establish an initial theory regarding the Second Order Asymptotical Regularization (SOAR) method for the stable approximate solution of ill-posed linear operator equations in Hilbert spaces, which are models for linear inverse problems with applications in the natural sciences, imaging and engineering. We show the regularizing properties of the new method, as well as the corresponding convergence rates. We prove that, under the appropriate source conditions and by using Morozov's conventional discrepancy principle, SOAR exhibits the same power-type convergence rate as the classical version of asymptotical regularization (Showalter's method). Moreover, we propose a new total energy discrepancy principle for choosing the terminating time of the dynamical solution from SOAR, which corresponds to the unique root of a monotonically non-increasing function and allows us to also show an order optimal convergence rate for SOAR. A damped symplectic iterative regularizing algorithm is developed for the realization of SOAR. Several numerical examples are given to show the accuracy and the acceleration affect of the proposed method. A comparison with other state-of-the-art methods are provided as well.

Keywords: Linear ill-posed problems, asymptotical regularization, second order method, convergence rate, source condition, index function, qualification, discrepancy principle

AMS Subject Classifications: 47A52;65J20;65F22;65R30

1. Introduction

We are interested in solving linear operator equations,

$$Ax = y, \tag{1}$$

where A is an injective and compact linear operator acting between two infinite dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. For simplicity, we denote by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ the inner products and norms, respectively, for both Hilbert spaces. Since A is injective, the operator equation (1) has a unique solution $x^\dagger \in \mathcal{X}$ for every y from range $\mathcal{R}(A)$ of the linear operator A. In this context, $\mathcal{R}(A)$ is assumed to be an infinite dimensional subspace of $\mathcal{Y}$.

Suppose that, instead of the exact right-hand side $y = Ax^\dagger$, we are given noisy data $y^\delta \in \mathcal{Y}$ obeying the deterministic noise model $\|y^\delta - y\| \le \delta$ with noise level $\delta > 0$. Since A is compact and $\dim(\mathcal{R}(A)) = \infty$, we have $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$ and the problem (1) is ill-posed. Therefore, regularization methods should be employed for obtaining stable approximate solutions.

*Corresponding author. Email: ye.zhang@mathematik.tu-chemnitz.de.

Loosely speaking, two groups of regularization methods exist: variational regularization methods and iterative regularization methods. Tikhonov regularization is certainly the most prominent variational regularization method (cf., e.g. [1]), while the Landweber iteration is the most famous iterative regularization approach (cf., e.g. [2, 3]). In this paper, our focus is on the latter, since from a computational viewpoint the iterative approach seems more attractive, especially for large-scale problems.

For the linear problem (1), the Landweber iteration is defined by

$$x_{k+1} = x_k + \Delta t A^*(y^\delta - Ax_k), \quad \Delta t \in (0, 2/\|A\|^2), \tag{2}$$

where A^* denotes the adjoint operator of A. We refer to [2, § 6.1] for the regularization property of the Landweber iteration. The continuous analogue of (2) can be considered as a first order evolution equation in Hilbert spaces

$$\dot{x}(t) + A^*Ax(t) = A^*y^\delta \tag{3}$$

if an artificial scalar time t is introduced, and $\Delta t \to 0$ in (2). Here and later on, we use Newton's conventions for the time derivatives. The formulation (3) is known as Showalter's method, or asymptotic regularization [4, 5]. The regularization property of (3) can be analyzed through a proper choice of the terminating time. Moreover, it has been shown that by using Runge-Kutta integrators, all of the properties of asymptotic regularization (3) carry over to its numerical realization [6].

From a computational view point, the Landweber iteration, as well as the steepest descent method and the minimal error method, is quite slow. Therefore, in practice accelerating strategies are usually used; see [3, 7] and references therein for details.

Over the last few decades, besides the first order iterative methods, there has been increasing evidence to show that the discrete second order iterative methods also enjoy remarkable acceleration properties for ill-posed inverse problems. The well-known methods are the Levenberg-Marquardt method [8], the iteratively regularized Gauss-Newton method [9], the ν-method [2, § 6.3], and the Nesterov acceleration scheme [10]. Recently, a more general second order iterative method – the two-point gradient method – has been developed, in [11]. In order to understand better the intrinsic properties of the discrete second order iterative regularization, we consider in this paper the continuous version

$$\begin{cases} \ddot{x}(t) + \eta\dot{x}(t) + A^*Ax(t) = A^*y^\delta, \\ x(0) = x_0, \quad \dot{x}(0) = \dot{x}_0. \end{cases} \tag{4}$$

of the second order iterative method in the form of an evolution equation, where $x_0 \in \mathcal{X}$ and $\dot{x}_0 \in \mathcal{X}$ are the prescribed initial data and $\eta > 0$ is a constant damping parameter.

From a physical viewpoint, the system (4) describes the motion of a heavy ball that rolls over the graph of the residual norm square functional $\Phi(x) = \|y^\delta - Ax\|^2$ and that keeps rolling under its own inertia until friction stops it at a critical point of $\Phi(x)$. This nonlinear oscillator with damping, which is called the *Heavy Ball with Friction* (HBF) system, has been considered by several authors from an optimization viewpoint, establishing different convergence results and identifying the circumstances under which the rate of convergence of HBF is better than the one of the first order methods; see [12–14]. Numerical algorithms based on (4) for solving some special problems, e.g. inverse source problems in partial differential equations,

large systems of linear equations, and the nonlinear Schrödinger problem, etc., can be found in [15–18]. The main goal of this paper is the intrinsic structure analysis of the theory of the second order iterative regularization and the development of new iterative regularization methods based on the framework (4).

The remainder of the paper is structured as follows: in Section 2, we extend the theory of general affine regularization schemes for solving linear ill-posed problems to a more general setting, adapted for the analysis of the second order model (4). Then, the existence and uniqueness of the second order flow (4), as well as some of its properties, are discussed in Section 3. Section 4 is devoted to the study of the regularization property of the dynamical solution to (4), while Section 5 presents the results about convergence rates under the assumption of conventional source conditions. In Section 6, based on the Störmer-Verlet method, we develop a novel iterative regularization method for the numerical implementation of the second order asymptotical regularization. Some numerical examples, as well as a comparison with four existing iterative regularization methods, are presented in Section 7. Finally, concluding remarks are given in Section 8.

2. General affine regularization methods

In this section, we consider general affine regularization schemes based on a family of pairs of piecewise continuous functions $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}_\alpha$ $(0 < \lambda \leq \|A\|^2)$ for regularization parameters $\alpha \in (0, \bar{\alpha}]$. Once a pair of generating functions $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}$ is chosen, the approximate solution to (1) can be given by the procedure

$$x_\alpha^\delta = (1 - A^*Ag_\alpha(A^*A))x_0 + \phi_\alpha(A^*A)\dot{x}_0 + g_\alpha(A^*A)A^*y^\delta. \tag{5}$$

Remark 1 *The affine regularization procedure defined by formula (5) is designed in particular for the second order evolution equation (4). If one sets* $(x_0, \dot{x}_0) = (0, 0)$*, the proposed regularization method coincides with the classical linear regularization schema for general linear ill-posed problems; see, e.g., [19]. However, as the numerical experiments in Section 7 will show, the initial data influence the behaviour of the regularized solutions obtained by (4). By finding an appropriate choice of the triple* $(x_0, \dot{x}_0, \eta)$*, the second order analog of the asymptotical regularization yields an accelerated procedure with approximate solutions of higher accuracy.*

To evaluate the regularization error $e(x^\dagger, \alpha, \delta) := \|x_\alpha^\delta - x^\dagger\|$ for the procedure (5) in combination with the the noise-free intermediate quantity

$$x_\alpha := (1 - A^*Ag_\alpha(A^*A))x_0 + \phi_\alpha(A^*A)\dot{x}_0 + g_\alpha(A^*A)A^*Ax^\dagger, \tag{6}$$

where evidently $e(x^\dagger, \alpha, \delta) \leq \|x_\alpha - x^\dagger\| + \|x_\alpha^\delta - x_\alpha\|$, we introduce the concepts of index functions and profile functions from [20] and [19], as follows:

Definition 1 *A real function* $\varphi : (0, \infty) \to (0, \infty)$ *is called an index function if it is continuous, strictly increasing and satisfies the condition* $\lim_{\lambda \to 0+} \varphi(\lambda) = 0$.

Definition 2 *An index function* f*, for which* $\|x_\alpha - x^\dagger\| \leq f(\alpha)$ $(\alpha \in (0, \bar{\alpha}])$ *holds, is called a profile function to* $x^\dagger$ *under the assumptions stated above.*

Having a profile function f estimating the noise-free error $\|x_\alpha - x^\dagger\|$ and taking into account that $\delta\|g_\alpha(A^*A)A^*\|$ is an upper bound for the noise-propagation error

$\|x_\alpha^\delta - x_\alpha\|$, which is independent of $x^\dagger$, we can estimate the total regularization error $e(x^\dagger, \alpha, \delta)$ as

$$e(x^\dagger, \alpha, \delta) \le f(\alpha) + \delta \|g_\alpha(A^*A)A^*\| \tag{7}$$

for all $\alpha \in (0, \bar{\alpha}]$. If we denote by

$$r_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda), \quad \lambda \in (0, \|A\|^2] \tag{8}$$

the bias function related to the major part of the regularization method g_α from (5), then $f(\alpha) = \|r_\alpha(A^*A)(x_0 - x^\dagger) + \phi_\alpha(A^*A)\dot{x}_0\|$ is evidently a profile function to $x^\dagger$ and we have

$$e(x^\dagger, \alpha, \delta) \le \|r_\alpha(A^*A)(x_0 - x^\dagger) + \phi_\alpha(A^*A)\dot{x}_0\| + \delta\|g_\alpha(A^*A)A^*\|. \tag{9}$$

PROPOSITION 1 *Assume that the pairs of functions $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}_\alpha$ are piecewise continuous in α and satisfy the following three conditions:*

(i) For any fixed $\lambda \in (0, \|A\|^2]$: $\lim_{\alpha\to 0} r_\alpha(\lambda) = 0$ and $\lim_{\alpha\to 0} \phi_\alpha(\lambda) = 0$.
(ii) Two constants γ_1 and γ_2 exist such that $|r_\alpha(\lambda)| \le \gamma_1$ and $|\phi_\alpha(\lambda)| \le \gamma_2$ hold for all $\alpha \in (0, \bar{\alpha}]$.
(iii) A constant γ_* exists such that $\sqrt{\lambda}|g_\alpha(\lambda)| \le \gamma_*/\sqrt{\alpha}$ for all $\alpha \in (0, \bar{\alpha}]$.

Then, if the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ is chosen so that

$$\lim_{\delta\to 0} \alpha = \lim_{\delta\to 0} \delta^2/\alpha = 0,$$

the approximate solution in (5) converges to the exact solution $x^\dagger$ as $\delta \to 0$.

Proof. From the properties (i) and (ii) of Proposition 1 we deduce for $\alpha \to 0$ point-wise convergence $r_\alpha(A^*A)x_1 \to 0$ and $\phi_\alpha(A^*A)x_2 \to 0$ for any $x_{1,2} \in \mathcal{X}$ (see, e.g., [2, Theorem 4.1]). Therefore, by the estimate (9) we can derive that

$$e(x^\dagger, \alpha, \delta) \le \|r_\alpha(A^*A)(x_0 - x^\dagger)\| + \|\phi_\alpha(A^*A)\dot{x}_0\| + \gamma_*\delta/\sqrt{\alpha} \to 0$$

as $\delta \to 0$.

■

Proposition 1 motivates us to call the procedure (5) a regularization method for the linear inverse problem (1) if the pair of functions $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}_\alpha$ satisfies the three requirements (i), (ii) and (iii).

EXAMPLE 1 *For the Landweber iteration (2) with the step size $\Delta t \in (0, 2/\|A\|^2)$, we have $\phi_\alpha(\lambda) = 0$ and $g_\alpha(\lambda) = (1 - (1 - \Delta t\lambda)^{\lfloor 1/\alpha \rfloor})/\lambda$. It is not difficult to show, e.g. in [19, 21], that $g_\alpha(\lambda)$ is a regularization method by Proposition 1 with constants $\gamma_2 = 0$, $\gamma_1 = 1$ and $\gamma_* = \sqrt{2\Delta t}$. Consider the continuous version of the Landweber iteration (3), i.e. Showalter's method. It is not difficult to show that $\phi_\alpha(\lambda) = 0$ and $g_\alpha(\lambda) = (1 - e^{-\lambda/\alpha})/\lambda$, and hence $r_\alpha(\lambda) = e^{-\lambda/\alpha}$. Obviously, $g_\alpha(\lambda)$ is a regularization method with $\gamma_2 = 0$, $\gamma_1 = 1$ and $\gamma_* = \theta$ by noting that $\sup_{\lambda\in\mathbf{R}_+} \sqrt{\lambda} g_\alpha(\lambda) = \theta\sqrt{\alpha}$ and $\theta = \sup_{\lambda\in\mathbf{R}_+} \sqrt{\lambda}(\lambda - e^{-\lambda}) \approx 0.6382$ [5].*

Note that the three requirements $(i) - (iii)$ in Proposition 1 are not enough to ensure rates of convergence for the regularized solutions. More precisely, for rates in the case of ill-posed problems, additional smoothness assumptions on $x^\dagger$

in correspondence with the forward operator A and the regularization method under consideration have to be fulfilled. This allows us to verify the specific profile functions $f(\alpha)$ in formula (7) that are specified for our second order method in formula (9). Once a profile function f is given, together with the property (*iii*) in Proposition 1, we obtain from the estimate (7) that

$$e(x^\dagger, \alpha, \delta) \leq f(\alpha) + \gamma_* \delta / \sqrt{\alpha} \quad \text{for all } \alpha \in (0, \bar{\alpha}]. \tag{10}$$

Moreover, if we consider the auxiliary index function

$$\Theta(\alpha) := \sqrt{\alpha} f(\alpha), \tag{11}$$

and choose the regularization parameter a priori as $\alpha_* = \Theta^{-1}(\delta)$, then we can easily see that

$$e(x^\dagger, \alpha_*, \delta) \leq (1 + \gamma_*) f(\Theta^{-1}(\delta)). \tag{12}$$

Hence, the convergence rate $f\left(\Theta^{-1}(\delta)\right)$ of the total regularization error as $\delta \to 0$ depends on the profile function f only, but for our suggested approach, f is a function of $x^\dagger, g_\alpha, \phi_\alpha, x_0, \dot{x}_0, A$ and on the damping parameter η.

In order to verify the profile function f in detail, it is of interest to consider how sensitive the regularization method is with respect to a priori smoothness assumptions. In this context, the concept of qualification can be exploited for answering this question: the higher its qualification, the more the method is capable of reacting to smoothness assumptions. Expressing the qualification by means of index functions ψ, the traditional concept of qualifications with monomials $\psi(\lambda) = \lambda^\kappa$ for $\kappa > 0$ from [5] (see also [2]) has been extended in [19, 20] to general index functions ψ. We adapt this generalized concept to our class of methods (5) in the following definition.

Definition 3 *Let ψ be an index function. A regularization method (5) for the linear operator equation (1) generated by the pair $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}$ $(0 < \lambda \leq \|A\|^2)$ is said to have the qualification ψ with constant $\gamma > 0$ if both inequalities*

$$\sup_{\lambda \in (0, \|A\|^2]} |r_\alpha(\lambda)| \psi(\lambda) \leq \gamma \psi(\alpha) \quad \textit{and} \quad \sup_{\lambda \in (0, \|A\|^2]} |\phi_\alpha(\lambda)| \psi(\lambda) \leq \gamma \psi(\alpha) \tag{13}$$

are satisfied for all $0 < \alpha \leq \|A\|^2$.

Remark 2 *Since the bias function of Showalter's method equals $r_\alpha(\lambda) = e^{-\lambda/\alpha}$ and $\phi_\alpha(\lambda) = 0$, set $\xi = \lambda$ in the following identity*

$$\sup_{0 \leq \xi < \infty} e^{-\xi/\alpha} \xi^p = (p/e)^p \, \alpha^p \tag{14}$$

to conclude that for all exponents $p > 0$ the monomials $\psi(\lambda) = \lambda^p$ are qualifications for Showalter's method. We will show that an analogue result also holds for the second order asymptotical regularization method - see Proposition 2 below - and will apply this fact to obtaining associated convergence rates.

3. Properties of the second order flow

We first prove the existence and uniqueness of strong global solutions of the second order equation (4). Then, we study the long-term behavior of the dynamical solution $x(t)$ of (4) and the residual norm functional $\|Ax(t) - y^\delta\|$.

DEFINITION 4 *$x : [0, +\infty) \to \mathcal{X}$ is a strong global solution of (4) with initial data $(x_0, \dot{x}_0)$ if $x(0) = x_0 \in \mathcal{X}, \dot{x}(0) = \dot{x}_0 \in \mathcal{X}$, and*

- *$x(\cdot), \dot{x}(\cdot) : [0, +\infty) \to \mathcal{X}$ are locally absolutely continuous [22],*
- *$\ddot{x}(t) + \eta\, \dot{x}(t) + A^*Ax(t) = A^*y^\delta$ holds for almost every $t \in [0, +\infty)$.*

THEOREM 1 *For any pair $(x_0, \dot{x}_0) \in \mathcal{X} \times \mathcal{X}$ there exists a unique strong global solution of the second order dynamical system (4).*

Proof. Denote by $\mathbf{z} = (x, \dot{x})^T$, and rewrite (4) as a first order differential equation

$$\dot{\mathbf{z}}(t) = B\mathbf{z}(t) + \mathbf{d}, \tag{15}$$

where $B = -[0, I; A^*A, \eta I]$, $\mathbf{d} = [0; A^*y^\delta]$ and I denotes the identity operator in $\mathcal{X}$. Since A is a bounded linear operator, both A^* and B are also bounded linear operators. Hence, by the Cauchy-Lipschitz-Picard theorem, the first-order autonomous system (15) has a unique global solution for the given initial data $(x_0, \dot{x}_0)$.

■

Now, we start to investigate the long-term behaviors of the dynamical solution and the residual norm functional. These properties will be used for the study of convergence rate in Section 5.

LEMMA 1 *Let $x(t)$ be the solution of (4). Then, $\dot{x}(\cdot) \in L^2([0, \infty), \mathcal{X})$ and $\dot{x}(t) \to 0$ as $t \to \infty$. Moreover, we have the following two limit relations*

$$\lim_{t\to\infty} \|Ax(t) - y^\delta\| \le \delta \tag{16}$$

and

$$\lim_{t\to\infty} \|Ax(t) - y^\delta\|^2 + \|\dot{x}(t)\|^2 = \inf_{x^* \in \mathcal{X}} \|Ax^* - y^\delta\|^2. \tag{17}$$

The proof of the above lemma uses the idea given in [14], and can be found in the Appendix A.1. If y^δ does not belong to the domain $\mathcal{D}(A^\dagger)$ of the Moore-Penrose inverse $A^\dagger$ of A, it is not difficult to show that there is a "blow-up" for the solution $x(t)$ of the dynamical system (4) in the sense that $\|x(t)\| \to \infty$ as $t \to \infty$. Contrarily, for $y^\delta \in \mathcal{D}(A^\dagger)$, i.e. if the noisy data y^δ satisfy the Picard condition, one can show more assertions concerning the long-term behaviour of the solution to the evolution equation (4), and we refer to Lemma 2 below for results in the case of noise-free data $y = Ax^\dagger$. In this work, for the inverse problem with noisy data, we are first and foremost interested in the case that $y^\delta \notin \mathcal{D}(A^\dagger)$ may occur, since the set $\mathcal{D}(A^\dagger)$ is of the first category and the chance to meet such an element is negligible.

At the end of this section, we show some properties of $x(t)$ of (4) with noise-free data.

LEMMA 2 *Let $x(t)$ be the solution of (4) with the exact right-hand side y as data. Then, in the case $\eta \geq \|A\|$, we have*

(i) $x(\cdot) \in L^\infty([0,\infty),\mathcal{X})$.
(ii) $\dot{x}(\cdot) \in L^\infty([0,\infty),\mathcal{X}) \cap L^2([0,\infty),\mathcal{X})$ *and* $\dot{x}(t) \to 0$ *as* $t \to \infty$.
(iii) $\ddot{x}(\cdot) \in L^\infty([0,\infty),\mathcal{X}) \cap L^2([0,\infty),\mathcal{X})$ *and* $\ddot{x}(t) \to 0$ *as* $t \to \infty$.
(iv) $\|Ax(t) - y\| = o(t^{-1/2})$ *as* $t \to \infty$.

The proof of Lemma 2 follows as a special case for $f(x) = \frac{1}{2}\|Ax - y\|^2$ in [22], and it is given in the Appendix A.2. The rate $\|Ax(t) - y\| = o(t^{-1/2})$ as $t \to \infty$ given in Lemma 2 for the second order evolution equation (4) should be compared with the corresponding result for the first order method, i.e. the gradient decent methods, where one only obtains $\|Ax(t) - y\| = \mathcal{O}(t^{-1/2})$ as $t \to \infty$. If we consider a discrete iterative method with the number k of iterations, assertion (iv) in Lemma 2 indicates that in comparison with gradient descent methods, the second order methods (4) need the same computational complexity for the number k of iterations, but can achieve a higher order $o(k^{-1/2})$ of accuracy for the objective functional as $k \to \infty$.

4. Convergence analysis for noisy data

This section is devoted to the verification of the pair $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}_\alpha$ of generator functions occurring in formula (6) associated with the second order equation problem (4) with the inexact right-hand side y^δ and the corresponding regularization properties.

Let $\{\sigma_j; u_j, v_j\}_{j=1}^\infty$ be the well-defined singular system for the compact and injective linear operator A, i.e. we have $Au_j = \sigma_j v_j$ and $A^* v_j = \sigma_j u_j$ with ordered singular values $\|A\| = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_j \geq \sigma_{j+1} \geq \cdots \to 0$ as $j \to \infty$. Since the eigenelements $\{u_j\}_{j=1}^\infty$ and $\{v_j\}_{j=1}^\infty$ form complete orthonormal systems in $\mathcal{X}$ and $\mathcal{Y}$, respectively, the equation in (4) is equivalent to

$$\langle \ddot{x}(t), u_j\rangle + \eta\langle \dot{x}(t), u_j\rangle + \sigma_j^2\langle x(t), u_j\rangle = \sigma_j\langle y^\delta, v_j\rangle, \quad j = 1, 2, \dots . \tag{18}$$

Using the decomposition $x(t) = \sum_j \xi_j(t) u_j$ under the basis $\{u_j\}_{j=1}^\infty$ in $\mathcal{X}$, we obtain

$$\ddot{\xi}_j(t) + \eta\dot{\xi}_j(t) + \sigma_j^2\xi_j = \sigma_j\langle y^\delta, v_j\rangle, \quad j = 1, 2, \dots . \tag{19}$$

In order to solve the above differential equation, we have to distinguish three different cases: (a) the overdamped case: $\eta > 2\|A\|$, (b) the underdamped case: there is an index j_0 such that $2\sigma_{j_0+1} < \eta < 2\sigma_{j_0}$, and (c) the critical damping case: an index j_0 exists such that $\eta = 2\sigma_{j_0}$. In this section, we discuss for simplicity the overdamped case only. The other two cases are studied similarly, and the corresponding details can be found in the Appendix B. We remark that all results that concluded in the overdamped case also hold for the other two cases, but with different value of positive constants $\gamma_{1,2}, \gamma_*$ in Proposition 1 and γ in Definition 3.

In the overdamped case, the characteristic equation of (19), possessing the form $\ddot{\xi}_j(t) + \eta\dot{\xi}_j(t) + \sigma_j^2\xi_j = 0$, which has two independent solutions $\xi_j^1 = e^{-\eta t/2}e^{\omega_j t/2}$ and $\xi_j^2 = e^{-\eta t/2}e^{-\omega_j t/2}$ for all $j = 1, 2, \dots$, where $\omega_j = \sqrt{\eta^2 - 4\sigma_j^2} > 0$. Hence, it is not

difficult to show that the general solution to (19) in the overdamped case is

$$\xi_j(t) = c_j^1 e^{-\eta t/2} e^{\omega_j t/2} + c_j^2 e^{-\eta t/2} e^{-\omega_j t/2} + \sigma_j^{-1}\langle y^\delta, v_j\rangle, \quad j = 1, 2, \dots . \tag{20}$$

Introducing the initial conditions in (4) to obtain a system for $\{c_j^1, c_j^2\}$ yields

$$\begin{cases} \sum_j \left(c_j^1 + c_j^2 + \sigma_j^{-1}\langle y^\delta, v_j\rangle \right) u_j = x^0, \\ \sum_j \left(\frac{\eta-\omega_j}{2} c_j^1 + \frac{\eta+\omega_j}{2} c_j^2 \right) u_j = \dot{x}^0, \end{cases} \tag{21}$$

or equivalently with the decomposition $x_0 = \sum_j \langle x_0, u_j\rangle u_j$ for all $j = 1, 2, ...$

$$\begin{cases} c_j^1 + c_j^2 + \sigma_j^{-1}\langle y^\delta, v_j\rangle = \langle x_0, u_j\rangle, \\ \frac{\eta-\omega_j}{2} c_j^1 + \frac{\eta+\omega_j}{2} c_j^2 = \langle \dot{x}_0, u_j\rangle, \end{cases} \tag{22}$$

which gives

$$\begin{cases} c_j^1 = \frac{\eta+\omega_j}{2\omega_j}\langle x_0, u_j\rangle - \frac{1}{\omega_j}\langle \dot{x}_0, u_j\rangle - \frac{\eta+\omega_j}{2\omega_j\sigma_j}\langle y^\delta, v_j\rangle, \\ c_j^2 = -\frac{\eta-\omega_j}{2\omega_j}\langle x_0, u_j\rangle + \frac{1}{\omega_j}\langle \dot{x}_0, u_j\rangle + \frac{\eta-\omega_j}{2\omega_j\sigma_j}\langle y^\delta, v_j\rangle. \end{cases} \qquad j = 1, 2, \dots . \tag{23}$$

By a combination of (23), the definition of ω_j and the decomposition of $x(t)$ we obtain

$$\begin{aligned} x(t) = &\sum_j \left(\frac{\eta+\sqrt{\eta^2-4\sigma_j^2}}{2\sqrt{\eta^2-4\sigma_j^2}} e^{-\frac{\eta-\sqrt{\eta^2-4\sigma_j^2}}{2}t} - \frac{\eta-\sqrt{\eta^2-4\sigma_j^2}}{2\sqrt{\eta^2-4\sigma_j^2}} e^{-\frac{\eta+\sqrt{\eta^2-4\sigma_j^2}}{2}t} \right) \langle x_0, u_j\rangle u_j \\ &- \sum_j \frac{1}{2\sqrt{\eta^2-4\sigma_j^2}} \left(e^{-\frac{\eta-\sqrt{\eta^2-4\sigma_j^2}}{2}t} - e^{-\frac{\eta+\sqrt{\eta^2-4\sigma_j^2}}{2}t} \right) \langle \dot{x}_0, u_j\rangle u_j \\ &+ \sum_j \frac{1 - \left(\frac{\eta+\sqrt{\eta^2-4\sigma_j^2}}{2\sqrt{\eta^2-4\sigma_j^2}} e^{-\frac{\eta-\sqrt{\eta^2-4\sigma_j^2}}{2}t} - \frac{\eta-\sqrt{\eta^2-4\sigma_j^2}}{2\sqrt{\eta^2-4\sigma_j^2}} e^{-\frac{\eta+\sqrt{\eta^2-4\sigma_j^2}}{2}t} \right)}{\sigma_j} \langle y^\delta, v_j\rangle u_j \\ =: &\, (1 - A^*Ag(t, A^*A))x_0 + \phi(t, A^*A)\dot{x}_0 + g(t, A^*A)A^*y^\delta, \end{aligned}$$

where

$$\begin{cases} g(t,\lambda) = \frac{1}{\lambda}\left(1 - \frac{\eta+\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}t} + \frac{\eta-\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta+\sqrt{\eta^2-4\lambda}}{2}t} \right), \\ \phi(t,\lambda) = -\frac{1}{2\sqrt{\eta^2-4\lambda}} \left(e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}t} - e^{-\frac{\eta+\sqrt{\eta^2-4\lambda}}{2}t} \right). \end{cases} \tag{24}$$

We find the form required for the generator functions in formula (6) if we set

$$g_\alpha(\lambda) := g(1/\alpha, \lambda) \quad \text{and} \quad \phi_\alpha(\lambda) := \phi(1/\alpha, \lambda). \tag{25}$$

Then the corresponding bias function $r_\alpha(\lambda) = 1 - \lambda g(1/\alpha, \lambda)$ is

$$r_\alpha(\lambda) = \frac{\eta + \sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}} - \frac{\eta - \sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta+\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}}. \tag{26}$$

THEOREM 2 *The functions $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}_\alpha$ in (25) based on (4) satisfy the conditions $(i)-(iii)$ of Proposition 1, which means that we consequently have a regularization method with the procedure (5) for the linear inverse problem (1).*

Proof. We check all of the three requirements in Proposition 1. The first condition obviously holds for $\phi_\alpha(\lambda)$ and $r_\alpha(\lambda)$, defined in (25) and (26) respectively.

The second condition can be obtained by using

$$\begin{cases} |r_\alpha(\lambda)| \le \frac{\eta+\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} = \frac{\eta}{2\sqrt{\eta^2-4\lambda}} + \frac{1}{2} \le \gamma_1 := \frac{\eta}{2\sqrt{\eta^2-4\|A\|^2}} + \frac{1}{2}, \\ |\phi_\alpha(\lambda)| \le \frac{1}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}t} \le \gamma_2 := \frac{\eta}{2\sqrt{\eta^2-4\|A\|^2}}. \end{cases} \tag{27}$$

It remains to bound γ_* in Proposition 1. By the inequality $1 - e^{-at} \le \sqrt{at}$ for $a > 0$, we obtain

$$1 - e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}} \le \sqrt{\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}} \frac{1}{\sqrt{\alpha}},$$

which implies that

$$\begin{aligned} \sqrt{\lambda}|g_\alpha(\lambda)| &= \frac{1}{\sqrt{\lambda}} \left(1 - \frac{\eta+\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}} + \frac{\eta-\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} e^{-\frac{\eta+\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}} \right) \\ &= \frac{1}{\sqrt{\lambda}} \frac{\eta+\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} \left(1 - e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}} \right) - \frac{1}{\sqrt{\lambda}} \frac{\eta-\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} \left(1 - e^{-\frac{\eta+\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}} \right) \\ &\le \frac{1}{\sqrt{\lambda}} \frac{\eta+\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}} \left(\sqrt{\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}} \frac{1}{\sqrt{\alpha}} \right) = \frac{1}{\sqrt{\eta^2-4\lambda}} \sqrt{\frac{\eta+\sqrt{\eta^2-4\lambda}}{2}} \frac{1}{\sqrt{\alpha}} \\ &\le \sqrt{\frac{\eta}{\eta^2-4\|A\|^2}} \frac{1}{\sqrt{\alpha}} \end{aligned}$$

Therefore, the third requirement in Proposition 1 holds for $g_\alpha(\lambda)$ with

$$\gamma_* = \sqrt{\eta/(\eta^2 - 4\|A\|^2)}. \tag{28}$$

Finally, by the proof above, we see that the upper bound $\bar{\alpha}$ for the affine regularization method with $\{g_\alpha(\lambda), \phi_\alpha(\lambda)\}_\alpha$ can be selected arbitrarily. ■

PROPOSITION 2 *For all exponents $p > 0$ the monomials $\psi(\lambda) = \lambda^p$ are qualifications with the constants*

$$\gamma = \left(\frac{p\eta}{e}\right)^p \left(\frac{\eta}{2\sqrt{\eta^2 - 4\|A\|^2}} + \frac{1}{2} \right) \tag{29}$$

for the second order asymptotical regularization method in the overdamped case.

Proof. Set $\xi = (\eta - \sqrt{\eta^2 - 4\lambda})/2$ in (14) and use the following inequality

$$\frac{\eta - \sqrt{\eta^2 - 4\lambda}}{2} = \frac{4\lambda}{2(\eta + \sqrt{\eta^2 - 4\lambda})} \ge \frac{\lambda}{\eta}$$

and the inequality (27), and we can derive that

$$\begin{aligned}\sup_{\lambda\in(0,\|A\|^2]}|r_\alpha(\lambda)|\psi(\lambda) &\le \sup_{\lambda\in(0,\|A\|^2]}\tfrac{\eta+\sqrt{\eta^2-4\lambda}}{2\sqrt{\eta^2-4\lambda}}e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}}\lambda^p\\ &\le \gamma_1\sup_{\lambda\in(0,\|A\|^2]}e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}}\lambda^p\\ &\le \gamma_1\eta^p\sup_{\lambda\in(0,\|A\|^2]}e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}}\left(\tfrac{\eta-\sqrt{\eta^2-4\lambda}}{2}\right)^p\\ &\le \gamma_1\eta^p\sup_{\xi\in(0,(\eta-\sqrt{\eta^2-4\|A\|^2})/2]}e^{-\xi/\alpha}\,(\xi)^p \le \gamma_1\eta^p\left(\tfrac{p}{e}\right)^p\alpha^p = \gamma\alpha^p.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\sup_{\lambda\in(0,\|A\|^2]}|\phi_\alpha(\lambda)|\psi(\lambda) &\le \sup_{\lambda\in(0,\|A\|^2]}\tfrac{1}{2\sqrt{\eta^2-4\lambda}}e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}}\lambda^p\\ &\le \gamma_2\sup_{\lambda\in(0,\|A\|^2]}e^{-\frac{\eta-\sqrt{\eta^2-4\lambda}}{2}\frac{1}{\alpha}}\lambda^p \le \gamma_2\eta^p\left(\tfrac{p}{e}\right)^p\alpha^p \le \gamma\alpha^p,\end{aligned}$$

which completes the proof. ■

The assertion of Theorem 2 and analogues to Proposition 2 can be found in the Appendix B for the other values of the constant $\eta > 0$ occurring as a parameter in the second order differential equation of problem (4). In particular, this means the underdamped case (b), as well as the critical damping case (c).

5. Convergence rates results

Under the general assumptions of the previous sections, the rate of convergence of $x(T) \to x^\dagger$ as $T \to \infty$ in the case of precise data, and of $x(T^*(\delta)) \to x^\dagger$ as $\delta \to 0$ in the case of noisy data, can be arbitrarily slow (cf. [23]) for solutions $x^\dagger$ which are not smooth enough. In order to prove convergence rates, some kind of smoothness assumptions imposed on the exact solution must be employed. Such smoothness assumptions can be expressed by range-type source conditions (cf., e.g., [2]), approximate source conditions (cf. [24]), and variational source conditions occurring in form of variational inequalities (cf. [25]). Now, range-type source conditions have the advantage that, in many cases, interpretations in the form of differentiability of the exact solution, boundary conditions, or similar properties are accessible. Hence, we focus in the following on the traditional range-type source conditions only. More precisely, we assume that an element $v_0 \in \mathcal{X}$ and numbers $p > 0$ and $\rho \ge 0$ exist such that

$$x_0 - x^\dagger = (A^*A)^p\, v_0 \quad \text{with} \quad \|v_0\| \le \rho. \tag{30}$$

Moreover, the initial data $\dot{x}_0$ is supposed to satisfy such source conditions as well, i.e.

$$\dot{x}_0 = (A^*A)^p\, v_1 \quad \text{with} \quad \|v_1\| \le \rho. \tag{31}$$

For the choice $\dot{x}_0 = 0$, the condition (31) is trivially satisfied. However, following the discussions in Sections 2 and 6, the regularized solutions essentially depend on

the value of $\dot{x}_0$. A good choice of $\dot{x}_0$ provides an acceleration of the regularization algorithm. In practice, one can choose a relatively small value of $\dot{x}_0$ to balance the source condition and the acceleration effect.

PROPOSITION 3 *Under the source conditions (30) and (31), $f(\alpha) = 2\gamma\rho\,\alpha^p$ is a profile function for the second order asymptotical regularization, where the constant γ is defined in (29).*

Proof. Combining the formulas (9), (30) and (31) yields

$$\begin{aligned}\|x(1/\alpha) - x^\dagger\| &\le \|r_\alpha(A^*A)(x_0 - x^\dagger)\| + \|\phi_\alpha(A^*A)\dot{x}_0\| \\ &\le \|r_\alpha(A^*A)\,(A^*A)^p\,v_0\| + \|\phi_\alpha(A^*A)\,(A^*A)^p\,v_1\| \\ &\le \rho \sup_{0\le\lambda\le\|A\|^2} |r_\alpha(\lambda)|\,\lambda^p + \rho \sup_{0\le\lambda\le\|A\|^2} |\phi_\alpha(\lambda)|\,\lambda^p \le 2\gamma\rho\,\alpha^p.\end{aligned}$$

This proves the proposition. ∎

THEOREM 3 *(A priori choice of the regularization parameter) If the terminating time T^* of the second order flow (4) is selected by the a priori parameter choice*

$$T^*(\delta) = c_0\rho^{2/(2p+1)}\,\delta^{-2/(2p+1)} \tag{32}$$

with the constant $c_0 = (2\gamma)^{2/(2p+1)}$, then we have the error estimate for $\delta \in (0, \delta_0]$

$$\|x(T^*) - x^\dagger\| \le c\rho^{1/(2p+1)}\,\delta^{2p/(2p+1)}, \tag{33}$$

where the constant $c = (1+\gamma_*)(2\gamma)^{1/(2p+1)}$ and $\delta_0 = 2\gamma\rho\eta^{2p+1}$.

Proof. By the discussion in Section 2, we choose the value of α_* such that $\delta = \Theta(\alpha_*) = \sqrt{\alpha_*}f(\alpha_*) = 2\gamma\rho\alpha_*^{p+1/2}$. By solving this equation we directly obtain $\alpha_* = (2\gamma)^{-2/(2p+1)}\ \rho^{-2/(2p+1)}\delta^{2/(2p+1)}$. Setting $T^* = 1/\alpha_*$ and using the estimate (12), this gives the relations (32) and

$$\begin{aligned}\|x(T^*) - x^\dagger\| = e(x^\dagger,\alpha,\delta) &\le (1+\gamma_*)f(\alpha_*) = (1+\gamma_*)2\gamma\rho(T^*)^{-p} \\ &= \left\{(1+\gamma_*)(2\gamma)^{1/(2p+1)}\right\}\rho^{1/(2p+1)}\delta^{2p/(2p+1)}.\end{aligned}$$

Finally, we use the inequality $\alpha_* \le \bar{\alpha} = \eta^2$ to get the bound δ_0 (the upper bound $\bar{\alpha} = \eta^2$ is required for the affine regularization (5) in both the underdamped and critical cases; see the appendix for details). ∎

In practice, the stopping rule in (32) is not realistic, since a good terminating time point T^* requires knowledge of ρ (a characteristic of unknown exact solution). Such knowledge, however, is not necessary in the case of a posteriori parameter choices. In the following two subsections, we consider two types of discrepancy principles for choosing the terminating time point a posteriori.

5.1. *Morozov's conventional discrepancy principle*

In our setting, Morozov's conventional discrepancy principle means searching for values $T > 0$ satisfying the equation

$$\chi(T) := \|Ax(T) - y^\delta\| - \tau\delta = 0, \tag{34}$$

where $\tau > \gamma_1 \geq 1$ is a constant, and the number γ_1 is defined in Proposition 1.

LEMMA 3 *If $\|Ax_0 - y^\delta\| > \tau\delta$, then the function $\chi(T)$ has at least one root.*

Proof. The continuity of $\chi(T)$ is obvious according to Theorem 1. On the other hand, from (16) and the assumption of the lemma, we conclude that

$$\lim_{T\to\infty} \chi(T) \leq (1-\tau)\delta < 0 \quad \text{and} \quad \chi(0) = \|Ax_0 - y^\delta\| - \tau\delta > 0,$$

which implies the existence of the root of $\chi(T)$. ■

THEOREM 4 *(A posteriori choice I of the regularization parameter) Suppose that $\|Ax_0 - y^\delta\| > \tau\delta$ and the source conditions (30) and (31) hold. If the terminating time T^* of the second order flow (4) is chosen according to the discrepancy principle (34), we have for any $\delta \in (0, \delta_0]$ and $p > 0$ the error estimates*

$$T^* \leq C_0 \rho^{2/(2p+1)} \delta^{-2/(2p+1)} \tag{35}$$

and

$$\|x(T^*) - x^\dagger\| \leq C_1 \delta^{2p/(2p+1)}, \tag{36}$$

where δ_0 is defined in the Theorem 3, $C_0 := (\tau - \gamma_1)^{-2/(2p+1)}(2\gamma)^{2/(2p+1)}$, and $C_1 := (\tau+\gamma_1)^{2p/(2p+1)} (\gamma_1+\gamma_2)^{1/(2p+1)} + \gamma_*(\tau-\gamma_1)^{-1/(2p+1)}(2\gamma)^{1/(2p+1)}$.

Proof. Using the moment inequality $\|B^p u\| \leq \|B^q u\|^{p/q}\|u\|^{1-p/q}$ and the source conditions (30)-(31), we deduce that

$$\begin{aligned}
&\|r_\alpha(A^*A^*)(x_0 - x^\dagger) + \phi_\alpha(A^*A^*)\dot{x}_0\| \\
&\quad = \|(A^*A)^{(p+1/2)} \left(r_\alpha(A^*A)v_0 + \phi_\alpha(A^*A)v_1\right)\| \\
&\quad \leq \|(A^*A)^{(p+1/2)} \left(r_\alpha(A^*A)v_0 + \phi_\alpha(A^*A)v_1\right)\|^{2p/(2p+1)} \\
&\qquad\qquad \cdot \|r_\alpha(A^*A)v_0 + \phi_\alpha(A^*A)v_1\|^{1/(2p+1)} \\
&\quad \leq \|Ar_\alpha(A^*A)(x_0 - x^\dagger) + A\phi_\alpha(A^*A^*)\dot{x}_0\|^{2p/(2p+1)} \\
&\qquad\qquad \cdot \left(\|r_\alpha(A^*A)v_0\| + \|\phi_\alpha(A^*A)v_1\|\right)^{1/(2p+1)}.
\end{aligned} \tag{37}$$

Since the terminating time T^* is chosen according to the discrepancy principle (34), we derive that

$$\begin{aligned}
\tau\delta &= \|Ax(T^*) - y^\delta\| \\
&= \left\|Ar_{1/T^*}(A^*A)(x_0 - x^\dagger) + A\phi_{1/T^*}(A^*A)\dot{x}_0 - r_{1/T^*}(A^*A)(y^\delta - y)\right\| \\
&\geq \|Ar_{1/T^*}(A^*A)(x_0 - x^\dagger) + A\phi_{1/T^*}(A^*A)\dot{x}_0\| - \|r_{1/T^*}(A^*A)(y^\delta - y)\|
\end{aligned} \tag{38}$$

Now we combine the estimates (37) and (38) to obtain, with the source conditions, that

$$\begin{aligned}
&\|r_\alpha(A^*A^*)(x_0 - x^\dagger) + \phi_\alpha(A^*A^*)\dot{x}_0\| \\
&\quad \leq \|Ar_\alpha(A^*A)(x_0 - x^\dagger) + A\phi_\alpha(A^*A^*)\dot{x}_0\|^{2p/(2p+1)} \\
&\qquad\qquad \cdot \left(\|r_\alpha(A^*A)v_0\| + \|\phi_\alpha(A^*A)v_1\|\right)^{1/(2p+1)} \\
&\quad \leq \left(\tau\delta + \|r_{1/T^*}(A^*A)(y^\delta - y)\|\right)^{2p/(2p+1)} \left((\gamma_1+\gamma_2)\rho\right)^{1/(2p+1)} \\
&\quad \leq c_1 \rho^{1/(2p+1)} \delta^{2p/(2p+1)}
\end{aligned} \tag{39}$$

where $c_1 := (\tau+\gamma_1)^{2p/(2p+1)}\,(\gamma_1+\gamma_2)^{1/(2p+1)}$.

On the other hand, in a similar fashion to (38), it is easy to show that

$$\begin{aligned}\tau\delta &\le \|Ar_{1/T^*}(A^*A)(x_0-x^\dagger)+A\phi_{1/T^*}(A^*A)\dot{x}_0\|+\|r_{1/T^*}(A^*A)(y^\delta-y)\|\\ &\le \|Ar_{1/T^*}(A^*A)(x_0-x^\dagger)+A\phi_{1/T^*}(A^*A)\dot{x}_0\|+\gamma_1\delta.\end{aligned}$$

If we combine the above inequality with the source conditions (30)-(31) and the qualification inequality (3), we obtain

$$\begin{aligned}(\tau-\gamma_1)\delta &\le \|Ar_{1/T^*}(A^*A)(x_0-x^\dagger)+A\phi_{1/T^*}(A^*A)\dot{x}_0\|\\ &\le \|(A^*A)^{p+1/2}r_{1/T^*}(A^*A)v_0\|+\|(A^*A)^{p+1/2}\phi_{1/T^*}(A^*A)v_1\|\le 2\rho\gamma(T^*)^{-(p+1/2)},\end{aligned}$$

which yields the estimate (35). Finally, using (35) and (39), we conclude that

$$\begin{aligned}\|x(T^*)-x^\dagger\| &\le \|r_\alpha(A^*A^*)(x_0-x^\dagger)+\phi_\alpha(A^*A^*)\dot{x}_0\|+\gamma_*\sqrt{T^*}\delta\\ &\le c_1\rho^{1/(2p+1)}\delta^{2p/(2p+1)}+\gamma_*\sqrt{C_0}\rho^{1/(2p+1)}\delta^{2p/(2p+1)}=C_1\rho^{1/(2p+1)}\delta^{2p/(2p+1)}.\end{aligned}$$

This completes the proof. ■

REMARK 3 *If the function $\chi(T)$ has more than one root, we recommend selecting T^* from the rule*

$$\chi(T^*)=0<\chi(T),\quad \forall T<T^*.$$

In other words, T^* is the first time point for which the size of the residual $\|Ax(T)-y^\delta\|$ has about the order of the data error. By Lemma 3 such T^* always exists.

It is easy to show that $\chi(T)$ is bounded by a decreasing function as the proof of Proposition 4 below will show. Roughly speaking, the trend of $\chi(T)$ is to be a decreasing function, where oscillations may occur, and we refer to Figure 5.1. On the other hand, one can anticipate that the more oscillations of the discrepancy function $\chi(T)$ occur, the smaller the damping parameter η is. This is an expected result due to the behaviour of damped Hamiltonian systems.

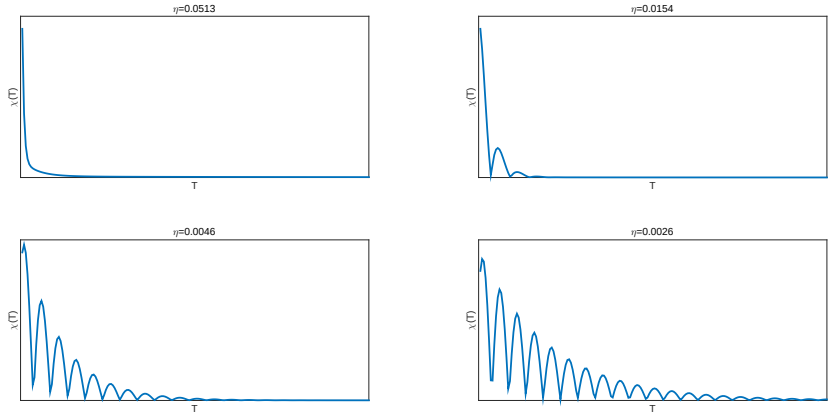

Figure 1. The behaviour of $\chi(T)$ from (34) with different damping parameters η.

5.2. *The total energy discrepancy principle*

For presenting a newly developed discrepancy principle, we introduce the total energy discrepancy function as follows:

$$\chi_{te}(T) := \|Ax(T) - y^\delta\|^2 + \|\dot{x}(T)\|^2 - \tau^2\delta^2, \tag{40}$$

where $\tau > \gamma_1$ as before.

PROPOSITION 4 *The function $\chi_{te}(T)$ is continuous and monotonically non-increasing. If $\|Ax_0 - y^\delta\|^2 + \|\dot{x}_0\|^2 > \tau^2\delta^2$, then $\chi_{te}(T) = 0$ has a unique solution.*

Proof. The continuity of $\chi_{te}(T)$ is obvious according to Theorem 1. The non-growth of $\chi_{te}(T)$ is straight-forward according to $\dot{\chi}_{te} = -2\eta\|\dot{x}\|^2$. Furthermore, from (16), (17) and the assumption of the proposition, we derive that

$$\lim_{T\to\infty} \chi_{te}(T) \leq \delta^2(1 - \tau^2) < 0, \tag{41}$$

and that, moreover, $\chi_{te}(0) = \|Ax_0 - y^\delta\|^2 + \|\dot{x}_0\|^2 - \tau^2\delta^2 > 0$. This implies the existence of roots for $\chi_{te}(T)$.

Finally, let us show that $\chi_{te}(T)$ has a unique solution. We prove this by contradiction. Since $\chi_{te}(T)$ is a non-increasing function, a number T_0 exists so that $\chi_{te}(T) = 0$ for $T \in [T_0, T_0 + \varepsilon]$ with some positive $\varepsilon > 0$. This means that $\dot{\chi}_{te}(T) = -2\eta\|\dot{x}\|^2 \equiv 0$ in $(T_0, T_0 + \varepsilon)$. Hence, $\ddot{x} \equiv 0$ in $(T_0, T_0 + \varepsilon)$. Using the equation (4) we conclude that for all $T > T_0$: $x(T) \equiv x(T_0)$. Since $\chi_{te}(T_0) = 0$, we obtain that $\chi_{te}(T) \equiv 0$ for $T > T_0$, which implies that $\lim_{T\to\infty} \chi_{te}(T) = 0$. This contradicts the fact in (41). ■

THEOREM 5 *(A posteriori choice II of the regularization parameter) Assume that $\|Ax_0 - y^\delta\|^2 + \|\dot{x}_0\|^2 > \tau^2\delta^2$ and a positive number δ_1 exists such that for all $\delta \in (0, \delta_1]$, the unique root T^* of $\chi_{te}(T)$ satisfies the inequality $\|Ax(T^*) - y^\delta\| \geq \tau_1\delta$, where $\tau_1 > \gamma_1$ is a constant, independent of δ. Then, under the source conditions (30) and (31), for any $\delta \in (0, \delta_0]$ and $p > 0$ we have the error estimates*

$$T^* \leq C_0\rho^{2/(2p+1)}\,\delta^{-2/(2p+1)}, \quad \|x(T^*) - x^\dagger\| \leq C_1\,\delta^{2p/(2p+1)}, \tag{42}$$

where δ_0 is defined in the Theorem 3, and constants C_0 and C_1 are the same as in Theorem 4.

Proof. The proof can be done along the lines and using the tools of the proof of Theorem 4. ■

In the simulation Section 7.1, we will computationally show that the assumptions occurring in the above theorem can happen in practice. Empirically, when the value of the initial velocity is not too small ($\|\dot{x}_0\| > 0$) or the noise is small enough ($\delta \ll 1$), the additional assumption $\|Ax(T^*) - y^\delta\| \geq \tau_1\delta$ in Theorem 5 always holds.

6. A novel iterative regularization method

Roughly speaking, the second order evolution equation (4) with an appropriate numerical discretization scheme for the artificial time variable yields a concrete

second order iterative method. Just as with the Runge-Kutta integrators [6] or the exponential integrators [26] for solving first order equations, the damped symplectic integrators are extremely attractive for solving second order equations, since the schemes are closely related to the canonical transformations [27], and the trajectories of the discretized second flow are usually more stable.

The simplest discretization scheme should be the Euler method. Denote by $v = \dot{x}$, and consider the following Euler iteration

$$\begin{cases} x^{k+1} = x^k + \Delta t_k v^k, \\ v^{k+1} = v^k + \Delta t_k \left(A^*(y^\delta - Ax^{k+1}) - \eta v^k \right), \\ x^0 = x_0, v^0 = \dot{x}_0. \end{cases} \tag{43}$$

By elementary calculations, scheme (43) expresses the form of following three-term semi-iterative method

$$x^{k+1} = x^k + \mu_k \left(x^k - x^{k-1} \right) + \omega_k A^*(y^\delta - Ax^k) \tag{44}$$

with a specially defined parameters $\omega_k = \Delta t_k$ and $\mu_k = 1 - \Delta t_k \eta$. It is well known that the semi-iterative method (44), equipped with an appropriate stopping rule, yields order optimal regularization method with (asymptotically) much fewer iterations than the calssical Landweber iteration [2, § 6.2].

In this paper, we develop a new iterative regularization method based on the Störmer-Verlet method, which also belongs to the family of symplectic integrators and takes the form

$$\begin{cases} v^{k+\frac{1}{2}} = v^k + \frac{\Delta t}{2} \left(A^*(y^\delta - Ax^k) - \eta v^{k+\frac{1}{2}} \right), \\ x^{k+1} = x^k + \Delta t v^{k+\frac{1}{2}}, \\ v^{k+1} = v^{k+\frac{1}{2}} + \frac{\Delta t}{2} \left(A^*(y^\delta - Ax^{k+1}) - \eta v^{k+\frac{1}{2}} \right), \\ x^0 = x_0, v^0 = \dot{x}_0. \end{cases} \tag{45}$$

PROPOSITION 5 *For any fixed damping parameter η, if the step size is chosen by*

$$\Delta t \leq \min \left(\sqrt{2}/\|A\|, 2/\eta \right), \tag{46}$$

then, the scheme (45) is convergent. Consequently, for any fixed T, there exists a pair of parameters $(k, \Delta t)$, satisfying $k\Delta t = T$ and the condition (46), such that $x^k = x(T) + \mathcal{O}(\Delta t^2)$ as $\Delta t \to 0$. Here x^k and $x(\cdot)$ are solutions to (45) and (4) respectively.

Proof. Denote by $\mathbf{z} = (x, v)^T$, and rewritten (45) as

$$\mathbf{z}^{k+1} = B\mathbf{z}^k + \frac{\Delta t}{2 + \Delta t \eta} \mathbf{b}, \tag{47}$$

where $\mathbf{b} = [\Delta t; 2I - \frac{\Delta t^2}{2} A^* A] A^* y_\delta$ and

$$B = \begin{bmatrix} I - \frac{\Delta t^2}{2+\Delta t\eta} A^* A & \frac{2\Delta t}{2+\Delta t\eta} I \\ -\frac{\Delta t}{2(2+\Delta t\eta)} \left(4I - \Delta t^2 A^* A \right) A^* A & \frac{2-\Delta t\eta}{2+\Delta t\eta} I - \frac{\Delta t^2}{2+\Delta t\eta} A^* A \end{bmatrix}.$$

By Taylor's theorem and the finite difference formula, it is not difficult to show

the consistency of the scheme (45). It is well known that boundedness implies the convergence of consistent schemes for any problem [28], hence, it suffices to show the boundedness of the scheme (45). The asymptotical behaviour $x^k = x(T) + \mathcal{O}(\Delta t^2)$ follows from the convergence result and the second order of the Störmer-Verlet method. Furthermore, a sufficient condition for the boundedness of the iterative algorithm (45) is that the operator B is non-expansive. Hence, it is necessary to prove that $\|B\|_2 \leq 1$.

Using the singular value decomposition, we have $A^*A = \Phi\Lambda\Phi^T$, where Φ is a unitary matrix and $\Lambda = \mathrm{diag}(\lambda_i)$, where $\lambda_i \geq 0, i = 1, ..., n$.

By the elementary calculations, we obtain that the eigenvalues of B equal

$$\mu_i^{\pm} = \frac{1}{2+\Delta t\eta}\left\{\left(2-\Delta t^2\lambda_i\right) \pm \sqrt{\left(2-\Delta t^2\lambda_i\right)^2 - \left(4-\Delta t^2\eta^2\right)}\right\}, i = 1, ..., n. \tag{48}$$

Denote by i_* the index of λ_{i_*}, corresponding to the maximal absolute value of $\mu_i^{\pm}$, i.e.

$$|\mu_{max}| = \frac{1}{2+\Delta t\eta}\max_{\pm}\left|\left(2-\Delta t^2\lambda_{i_*}\right) \pm \sqrt{\left(2-\Delta t^2\lambda_{i_*}\right)^2 - \left(4-\Delta t^2\eta^2\right)}\right|.$$

There are three possible cases here: the overdamped case ($\left(2-\Delta t^2\lambda_{i_*}\right)^2 > 4-\Delta t^2\eta^2$), the underdamped case ($\left(2-\Delta t^2\lambda_{i_*}\right)^2 < 4-\Delta t^2\eta^2$), and the critical damping case ($\left(2-\Delta t^2\lambda_{i_*}\right)^2 = 4-\Delta t^2\eta^2$).

Let us consider these cases respectively. For the chosen time step size Δt in (46), we have $2-\Delta t^2\lambda_{i_*} \geq 0$. Therefore, for the overdamped case,

$$|\mu_{max}| = \frac{1}{2+\Delta t\eta}\left\{\left(2-\Delta t^2\lambda_{i_*}\right) + \sqrt{\left(2-\Delta t^2\lambda_{i_*}\right)^2 - \left(4-\Delta t^2\eta^2\right)}\right\}.$$

Define $2-\Delta t^2\lambda_{i_*} = a\sqrt{4-\Delta t^2\eta^2}$ with $a > 1$ (by the condition (46). It holds that $4-\Delta t^2\eta^2 \geq 0$), and we have

$$|\mu_{max}| = \frac{(a+\sqrt{a^2-1})\sqrt{4-\Delta t^2\eta^2}}{2+\Delta t\eta} = \frac{(a+\sqrt{a^2-1})(2-\Delta t^2\lambda_{i_*})}{a(2+\Delta t\eta)}. \tag{49}$$

Note that

$$\Delta t\eta = \sqrt{4-\left(\frac{2-\Delta t^2\lambda_{i_*}}{a}\right)^2}. \tag{50}$$

Combine (49) and (50) to obtain

$$|\mu_{max}| = \frac{(a+\sqrt{a^2-1})(2-\Delta t^2\lambda_{i_*})}{2a+\sqrt{4a^2-\left(2-\Delta t^2\lambda_{i_*}\right)^2}} = \frac{(a+\sqrt{a^2-1})(1-\frac{\Delta t^2\lambda_{i_*}}{2})}{a+\sqrt{a^2-\left(1-\frac{\Delta t^2\lambda_{i_*}}{2}\right)^2}} \leq 1-\frac{\Delta t^2\lambda_{i_*}}{2} \leq 1.$$

Now, consider the underdamped case. In this case, the complex eigenvalue μ_{max}

satisfies

$$|\mu_{max}|^2 = \frac{1}{(2+\Delta t\eta)^2}\left\{\left(2-\Delta t^2\lambda_{i_*}\right)^2 + \left[\left(4-\Delta t^2\eta^2\right) - \left(2-\Delta t^2\lambda_{i_*}\right)^2\right]\right\},$$

which implies that

$$|\mu_{max}|^2 = \frac{4-\Delta t^2\eta^2}{(2+\Delta t\eta)^2} = \frac{2-\Delta t\eta}{2+\Delta t\eta} < 1.$$

Finally, consider the critical damping case. Similarly, we have $|\mu_{max}| = \sqrt{\frac{2-\Delta t\eta}{2+\Delta t\eta}} < 1$, which yields the desired result.

■

At the end of this second, we show the convergence rate of the scheme (45).

THEOREM 6 *Under the assumptions of Theorem 4 or 5, if the time step size is chosen by $\Delta t = C_t\delta^{p/(2p+1)}$, then for any $\delta \in (0, \delta_s]$ and $p > 0$ we have the convergence rate*

$$\|x^{k^*} - x^\dagger\| = \mathcal{O}(\delta^{2p/(2p+1)}), \tag{51}$$

where $k^* = T^*/\Delta t$, $\delta_s = \min\left\{(\sqrt{2}/C_t)^{2+1/p}\|A\|^{-2-1/p}, (\sqrt{2}/C_t)^{2+1/p}\eta^{-2-1/p}, \delta_0\right\}$, $C_t = (T^*/\delta^{p/(2p+1)})/\lfloor T^*/\delta^{p/(2p+1)}\rfloor$, and T^* is the root of (34) or (40). Here, $\lfloor\cdot\rfloor$ denotes the standard floor function and δ_0 is defined in the Theorem 3.

Proof. It follows from Proposition 5 that the choice $\Delta t = C_t\delta^{p/(2p+1)}$ yields

$$\|x^{k^*} - x(T^*)\| \le C\delta^{2p/(2p+1)}$$

with some constant C for all $\delta \in (0, \delta_s]$.

Combine the above inequality and the results in theorems 4 and 5 to obtain

$$\|x^{k^*} - x^\dagger\| \le \|x^k - x(T^*)\| + \|x(T^*) - x^\dagger\| \le C_r\,\delta^{2p/(2p+1)}$$

for some constant C_r. This gives the estimate (51). ■

7. Numerical simulations

In this section, we present some numerical results for the following integral equation

$$Ax(s) := \int_0^1 K(s,t)x(t)dt = y(s), \quad K(s,t) = s(1-t)\chi_{s\le t} + t(1-s)\chi_{s>t}. \tag{52}$$

If we choose $\mathcal{X} = \mathcal{Y} = L^2[0,1]$, the operator A is compact, selfadjoint and injective. It is well known that the integral equation (52) has a solution $x = -y''$ if $y \in H^2[0,1] \cap H_0^1[0,1]$. Moreover, the operator A has the eigensystem $Au_j = \sigma_j u_j$, where $\sigma_j = (j\pi)^{-2}$ and $u_j(t) = \sqrt{2}\sin(j\pi t)$. Furthermore, using the interpolation theory (see e.g. [29]) it is not difficult to show that for $4p - 1/2 \notin \mathbf{N}$

$$R((A^*A)^p) = \left\{x \in H^{4p}[0,1] :\ x^{2l}(0) = x^{2l}(1) = 0,\ l = 0, 1, ..., \lfloor 2p - 1/4\rfloor\right\}.$$

In general, a regularization procedure becomes numerically feasible only after an appropriate discretization. Here, we apply the linear finite elements to solve (52). Let $\mathcal{Y}_n$ be the finite element space of piecewise linear functions on a uniform grid with step size $1/(n-1)$. Denote by P_n the orthogonal projection operator acting from $\mathcal{Y}$ into $\mathcal{Y}_n$. Define $A_n := P_n A$ and $\mathcal{X}_n := A_n^* \mathcal{Y}_n$. Let $\{\phi_j\}_{j=1}^n$ be a basis of the finite element space $\mathcal{Y}_n$, then, instead of the original problem (52), we solve the following system of linear equations

$$A_n x_n = y_n, \tag{53}$$

where $[A_n]_{ij} = \int_0^1 \left(\int_0^1 k(s,t)\phi_i(s)ds \right) \phi_j(t)dt$ and $[y_n]_j = \int_0^1 y(t)\phi_j(t)dt$.

As shown in [2], the finite dimensional projection error $\epsilon_n := \|(I - P_n)A\|$ plays an important role in the convergence rates analysis. For the compact operator A, $\epsilon_n \to 0$ as $n \to \infty$. Moreover, if the noise level $\delta \to 0$ slowly enough as $n \to \infty$, the quality ϵ_n has no influence and we obtain the same convergence rates as in theorems 4 and 5.

Uniformly distributed noises with the magnitude δ' are added to the discretized exact right-hand side:

$$[y_n^\delta]_j := \left[1 + \delta' \cdot (2\text{Rand}(x) - 1)\right] \cdot [y_n]_j, \quad j = 1, ..., n, \tag{54}$$

where Rand(x) returns a pseudo-random value drawn from a uniform distribution on [0,1]. The noise level of measurement data is calculated by $\delta = \|y_n^\delta - y_n\|_2$, where $\|\cdot\|_2$ denotes the standard vector norm in $\mathbf{R}^n$.

To assess the accuracy of the approximate solutions, we define the L^2-norm relative error for an approximate solution $x_n^{k^*}$ ($k^* = \lfloor T^*/\Delta t \rfloor$):

$$\text{L2Err} := \|x_n^{k^*} - x^\dagger\|_{L^2[0,1]} / \|x^\dagger\|_{L^2[0,1]},$$

where $x^\dagger$ is the exact solution to the corresponding model problem.

7.1. *Influence of parameters*

The purpose of this subsection is to explore the dependence of the solution accuracy and the convergence rate on the initial data $(x_0, \dot{x}_0)$, damping parameter η and the discrepancy functions χ and χ_{te}, and thus to give a guide on the choices of them in practice.

In this subsection, we solve integral equation (52) with the exact right-hand side $y = s(1-s)$. Then, the exact solution $x^\dagger = 2$, and $x^\dagger \in R((A^*A)^p)$ for all $p < 1/8$. Denote by "DP" and "TEDP" the newly developed iterative scheme (45) equipped with the Morozov's conventional discrepancy function $\chi(T)$ and the total energy discrepancy functions $\chi_{te}(T)$ respectively.

The results about the influence of the solution accuracy (L2Err) and the convergence rate (iteration numbers $k^*(\delta)$) on the initial data $(x_0, \dot{x}_0)$ are given in Tables 1 and 2 respectively. As we can see, both the initial data x_0 and $\dot{x}_0$ influence the solution accuracy as well as the convergence rate. Moreover, when the value of the damping parameter is not too small (see Tab. 2 and Tab. 3) the results (both solution accuracy and convergence rate) by the methods "DP" and "TEDP" almost coincide with each other. This result verifies the rationality of the assumption in Theorem 5.

Table 1. The dependence of the solution accuracy and the convergence rate on the initial data x_0. $\Delta t = 19.4946, \eta = 2.5648 \times 10^{-4}, \dot{x}_0 = 0, \tau = 2, p = 0.1125, \tau_{te} = 1.1 \times \delta^{4p/(4p+1)}$.

δ (δ')	x_0	DP		TEDP	
		$k^*(\delta)$	L2Err	$k^*(\delta)$	L2Err
7.1191e-04 (1e-03)	0	1295	0.2214	2033	0.2894
7.1337e-04 (1e-03)	-1	1354	0.2284	2210	0.2825
7.0838e-04 (1e-03)	1	1108	0.1965	1777	0.1414
7.1602e-05 (1e-04)	0	2397	0.0688	2958	0.0653
8.0523e-05 (1e-04)	-1	2525	0.0983	3076	0.0920
7.8903e-05 (1e-04)	1	1600	0.0449	2643	0.0431

Table 2. The dependence of the solution accuracy and the convergence rate on the initial data $\dot{x}_0$. $\Delta t = 19.4946, \eta = 0.0154, x_0 = 1, \tau = 2, p = 0.1125, \tau_{te} = 1.1 \times \delta^{4p/(4p+1)}$.

δ (δ')	$\dot{x}_0$	DP		TEDP	
		$k^*(\delta)$	L2Err	$k^*(\delta)$	L2Err
6.3541e-04 (1e-03)	0	155	0.1062	156	0.1061
7.3673e-04 (1e-03)	-0.01	374	0.1576	376	0.1574
7.1484e-04 (1e-03)	0.01	39	0.0443	39	0.0443
8.0763e-05 (1e-04)	0	4559	0.0675	4561	0.0675
7.3174e-05 (1e-04)	-0.01	12466	0.0953	12467	0.0953
7.6200e-05 (1e-04)	0.01	982	0.0293	983	0.0293

In Table 3, we displayed the numerical results with different value of damping parameters η. With the appropriate choice of the damping parameter, say $\eta = 2.5648 \times 10^{-3}$ in our example, the second order asymptotical regularization not only gives the most accurate result, but exhibits an acceleration affect. The critical value of the damping parameter, say $\eta = 2/\Delta t$, also provides an accurate result. But it requires a few more steps. The influence of the damping parameter on the residual functional can be found in Figure 5.1. It shows that at the same time point, the larger the damping parameter, the smaller the residual norm functional.

Table 3. The dependence of the solution accuracy and the convergence rate on the damping parameter η. $\Delta t = 19.4946, x_0 = 1, \dot{x}_0 = 0, \tau = 2, p = 0.1125, \tau_{te} = 1.1 \times \delta^{4p/(4p+1)}$.

δ (δ')	η	DP		TEDP	
		$k^*(\delta)$	L2Err	$k^*(\delta)$	L2Err
8.7715e-04 (1e-03)	2.5648e-05	6943	0.8319	17393	0.7744
7.3673e-04 (1e-03)	2.5648e-04	1108	0.1310	1728	0.1080
7.1484e-04 (1e-03)	2.5648e-03	124	0.0901	183	0.0870
8.1011e-05 (1e-04)	2.5648e-02	205	0.1101	206	0.1100
7.3174e-05 (1e-04)	0.0513	398	0.1108	398	0.1108
7.2721e-05 (1e-04)	0.1026 (= $2/\Delta t$)	848	0.1096	848	0.1096

7.2. *Comparison with other methods*

In order to demonstrate the advantages of our algorithm over the traditional approaches, we solve the same problems by four famous iterative regularization methods – the Landweber method, the Nesterov's method, the ν-method and the conjugate gradient method for the normal equation (CGNE, cf., e.g. [30]). The Landweber method is given in (2), while Nesterov's method is defined as [31]

$$\begin{cases} z^k = x^k + \frac{k-1}{k+\alpha-1}(x^k - x^{k-1}), \\ x^{k+1} = z^k + \Delta t A^*(y^\delta - Az^k), \end{cases} \tag{55}$$

where $\alpha > 3$ (we choose $\alpha = 3.1$ in our simulations). Moreover, we select the Chebyshev method as our special ν-method, i.e., $\nu = 1/2$ [2, § 6.3]. For all of four traditional iterative regularization methods, we use the Morozov's conventional discrepancy principle as the stopping rule.

We consider the following two different right-hand sides for the integral equation (52).

- **Example 1**: $y(s) = s(1-s)$. Then, $x^\dagger = 2$, and $x^\dagger \in R((A^*A)^p)$ for all $p < 1/8$. This example uses the discretization size $n = 400$. Other parameters are: $\Delta t = 19.4946, \eta = 2.5648 \times 10^{-4}, x_0 = 1, \dot{x}_0 = 0, \tau = 2, p = 0.1125, \tau_{te} = 1.1 \times \delta^{4p/(4p+1)}$.
- **Example 2**: $y(s) = s^4(1-s)^3$. Then, $x^\dagger = -6t^2(1-t)(2-8t+7t^2)$, and $x^\dagger \in R((A^*A)^p)$ for all $p < 5/8$. This example uses the discretization size $n = 400$. Other parameters are: $\Delta t = 19.4946, \eta = 0.0051, x_0 = 0, \dot{x}_0 = 0, \tau = 2, p = 0.5625, \tau_{te} = 1.1 \times \delta^{4p/(4p+1)}$.

The results of the simulation are presented in Table 4, where we can conclude that, in general, the second order asymptotical regularization need less iteration and CPU time, and offers a more accurate regularization solution. Moreover, with respect to the Morozov's conventional discrepancy principle, the newly developed total energy discrepancy principle provides more accurate results in most cases. Concerning the number of iterations, the CGNE method performed much better than all of other methods. However, the accuracy of the CGNE method is much worse than other methods (including DP and TEDP), since the step size of CGNE is too large to capture the optimal point and the semi-convergence effect disturbs the iteration rather early. Note that we set a maximal iteration number $k_{max} = 400,000$ in all of our simulations.

Table 4. Comparisons with the Landweber method, the Nesterov's method, the Chebyshev method, and the CGNE method.

δ	DP			TEDP			Landweber		
	$k^*(\delta)$	CPU (s)	L2Err	$k^*(\delta)$	CPU (s)	L2Err	$k^*(\delta)$	CPU (s)	L2Err
Example 1									
1.0400e-2	70	0.2023	0.1494	70	0.1995	0.1494	20438	52.6428	0.2639
1.0380e-3	2872	8.9239	0.0945	2871	1.9105	0.0945	k_{max}	1.6899e3	0.1807
1.0445e-4	56246	177.3387	0.0597	56246	177.9603	0.0473	k_{max}	1.8048e3	0.1807
Example 2									
1.0761e-2	15	0.0554	0.3676	25	0.0753	0.1987	1457	6.3672	0.7032
1.0703e-3	93	0.2488	0.0637	132	0.3567	0.0581	23790	67.6082	0.1767
1.1006e-4	453	1.4931	0.0195	531	1.8000	0.0184	187188	679.8274	0.0509
δ	Nesterov			Chebyshev			CGNE		
	$k^*(\delta)$	CPU (s)	L2Err	$k^*(\delta)$	CPU (s)	L2Err	$k^*(\delta)$	CPU (s)	L2Err
Example 1									
1.0400e-2	419	1.1384	0.2590	264	0.7378	0.2553	6	0.0351	0.2213
1.0380e-3	2813	8.5986	0.1600	2229	7.6512	0.1496	18	0.0904	0.1383
1.0445e-4	16642	60.1179	0.1025	17443	52.2056	0.0897	39	0.2013	0.0894
Example 2									
1.0761e-2	102	0.3018	0.7043	62	0.1768	0.7102	6	0.0148	0.4835
1.0703e-3	416	1.1732	0.1676	415	1.1908	0.1190	12	0.0261	0.1514
1.1006e-4	1805	6.0793	0.0280	2226	7.6967	0.0196	15	0.0309	0.0447

8. Conclusion and outlook

In this paper, we have investigated the method of second order asymptotical regularization (SOAR) for solving the ill-posed linear inverse problem $Ax = y$ with the compact operator A mapping between infinite dimensional Hilbert spaces. Instead of y, we are given noisy data y^δ obeying the inequality $\|y^\delta - y\| \leq \delta$. We have shown regularization properties for the dynamical solution of the second order equation (4). Moreover, by using Morozov's conventional discrepancy principle on the one hand and a newly developed total energy discrepancy principle on the other hand, we have proven the order optimality of SOAR. Furthermore, based the framework of SOAR, by using the Störmer-Verlet method, we have derived a novel iterative regularization algorithm. The convergence property of the proposed numerical algorithm is proven as well. Numerical experiments in Section 7 show that, in comparison with conventional iterative regularization methods, SOAR is a faster regularization method for solving linear inverse problems with high levels of accuracy.

Various numerical results show that the damping parameter η in the second order equation (4) plays a prominent role in regularization and acceleration. Therefore, how to choose an optimal damping parameter should be studied in the future. Moreover, using the results of the nonlinear Landweber iteration, it will be possible to develop a theory of second order asymptotical regularization for wide classes of nonlinear ill-posed problems. Furthermore, it could be very interesting to investigate the case with the dynamical damping parameter $\eta = \eta(t)$. For instance, the second order equation (4) with $\eta = r/t$ ($r \geq 3$) presents the continuous version of Nesterov's scheme [32], and the discretization of (4) with

$$\begin{cases} \eta_k = \frac{(k+2\nu-1)(2k+4\nu-1)(2k+2\nu-3)-(k-1)(2k-3)(3k+3\nu-1)}{4(2k+2\nu-3)(2k+2\nu-1)(k+\nu-1)}, \\ \Delta t_k = 4\frac{(2k+2\nu-1)(k+\nu-1)}{(k+2\nu-1)(2k+4\nu-1)}, \end{cases}$$

yields the well-known ν-methods [2, § 6.3]. Even in the linear case (1), to the best of our knowledge, it is not quite clear whether Nesterov's approach equipped with a posteriori choice of the regularization parameter is an accelerated regularization method for solving ill-posed inverse problems. In our opinion, however, the second order asymptotical regularization can be a candidate for the systematic analysis of general second order regularization methods.

9. Acknowledgement

The work of Y. Zhang is supported by the Alexander von Humboldt foundation through a postdoctoral researcher fellowship, and the work of B. Hofmann is supported by the German Research Foundation (DFG-grant HO 1454/12-1).

References

[1] Tikhonov A, Leonov A, Yagola A. Nonlinear ill-posed problems, volumes i and ii. London: Chapman and Hall; 1998.
[2] Engl H, Hanke M, Neubauer A. Regularization of inverse problems. New York: Springer; 1996.
[3] Kaltenbacher B, Neubauer A, Scherzer O. Iterative regularization methods for nonlinear ill-posed problems. Berlin: Walter de Gruyter GmbH & Co. KG; 2008.

[4] Tautenhahn U. On the asymptotical regularization of nonlinear ill-posed problems. Inverse Problems. 1994;10:1405–1418.
[5] Vainikko G, Veretennikov A. Iteration procedures in ill-posed problems. Moscow: Nauka (In Russian); 1986.
[6] Rieder A. Runge-Kutta integrators yield optimal regularization schemes. Inverse Problems. 2005;21:453–471.
[7] Neubauer A. On Landweber iteration for nonlinear ill-posed problems in Hilbert scales. Numerische Mathematik. 2000;85:309–328.
[8] Jin Q. On a regularized Levenberg-Marquardt method for solving nonlinear inverse problems. Numerische Mathematik. 2010;115:229–259.
[9] Jin Q. On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems. Numerische Mathematik. 2009;111:509–558.
[10] Neubauer A. On Nesterov acceleration for Landweber iteration of linear ill-posed problems. Journal of Inverse and Ill-Posed Problems. 2017;25:381–390.
[11] Hubmer S, Ramlau R. Convergence analysis of a two-point gradient method for nonlinear ill-posed problems. Inverse Problems. 2017;33:095004.
[12] Alvarez F. On the minimizing property of a second-order dissipative system in Hilbert spaces. SIAM J Control Optim. 2000;38:1102–1119.
[13] Alvarez F, Attouch H, Bolte J, Redont P. A second-order gradient-like dissipative dynamical system with hessian-driven damping. application to optimization and mechanics. J Math Pures Appl. 2002;81:747–779.
[14] Attouch H, Goudou X, Redont P. The heavy ball with friction method. i. the continuous dynamical system. Comm Contemp Math. 2000;2:1–34.
[15] Zhang Y, Gong R, Cheng X, Gulliksson M. A dynamical regularization algorithm for solving inverse source problems of elliptic partial differential equations. Inverse Problems. 2018;34:065001.
[16] Edvardsson S, Neuman M, Edström P, Olin H. Solving equations through particle dynamics. Comput Phys Commun. 2015;197:169–181.
[17] Edvardsson S, Gulliksson M, Persson J. The dynamical functional particle method: an approach for boundary value problems. J Appl Mech. 2012;79:021012.
[18] Sandin P, Ögren M, Gulliksson M. Numerical solution of the stationary multicomponent nonlinear schrodinger equation with a constraint on the angular momentum. Phys Rev E. 2016;93:033301.
[19] Hofmann B, Mathé P. Analysis of profile functions for general linear regularization methods. SIAM J Numer Anal. 2007;45:1122–1141.
[20] Mathé P, Pereverzev SV. Geometry of linear ill-posed problems in variable Hilbert scales. Inverse Problems. 2003;19(3):789–803.
[21] Mathé P. Saturation of regularization methods for linear ill-posed problems in Hilbert spaces. SIAM J Numer Anal. 2006;42:968–973.
[22] Boţ R, Csetnek E. Second order forward-backward dynamical systems for monotone inclusion problems. SIAM Journal on Control and Optimization. 2016;54:1423–1443.
[23] Schock E. Approximate solution of ill-posed equations: arbitrarily slow convergence vs. superconvergence. Constructive methods for the practical treatment of integral equations. 1985;73:234–243.
[24] Hein T, Hofmann B. Approximate source conditions for nonlinear ill-posed problems – chances and limitations. Inverse Problems. 2009;25:035003.
[25] Hofmann B, Kaltenbacher B, Poschl C, Scherzer O. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. Inverse Problems. 2007;23:987–1010.
[26] Hochbruck M, Lubich C, Selhofer H. Exponential integrators for large systems of differential equations. SIAM J Sci Comput. 1998;19:1152–1174.
[27] Hairer E, Wanner G, Lubich C. Geometric numerical integration: Structure-preserving algorithms for ordinary differential equations (second edition). New York: Springer; 2006.
[28] Tadmor E. A review of numerical methods for nonlinear partial differential equations. Bulletin of the American Mathematical Society. 2012;49(4):507–554.

[29] Lions J, Magenes E. Non-homogeneous boundary value problems and applications, volumes i. Berlin: Springer; 1972.
[30] Hanke M. Conjugate gradient type methods for ill-posed problems. New York: John Wiley & Sons, Inc.; 1995.
[31] Nesterov Y. A method of solving a convex programming problem with convergence rate. Soviet Mathematics Doklady. 1983;27:372–376.
[32] Su W, Boyd S, Candes E. A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. Journal of Machine Learning Research. 2016; 17:1–43.

Appendix A. Proofs in Section 3

A.1. *Proof of Lemma 1*

Define the Lyapunov function of (4) by $\mathcal{E}(T) = \|Ax(T) - y^\delta\|^2 + \|\dot{x}(T)\|^2$. It is not difficult to show that

$$\dot{\mathcal{E}}(t) = -2\eta\|\dot{x}(t)\|^2 \tag{A1}$$

by looking at the equation (4) and the differentiation of the energy function $\dot{\mathcal{E}}(t) = 2\langle\dot{x}(t), \ddot{x}(t) - A^*(y^\delta - Ax(t))\rangle$. Hence, $\mathcal{E}(t)$ is non-increasing, and consequently, $\|\dot{x}(t)\|^2 \leq \mathcal{E}(0)$. Therefore, $\dot{x}(\cdot)$ is uniform bounded. Integrating both sides in (A1), we obtain

$$\int_0^\infty \|\dot{x}(t)\|^2 dt \leq \mathcal{E}(0)/(2\eta) < \infty,$$

which yields $\dot{x}(\cdot) \in L^2([0,\infty), \mathcal{X})$.

Now, let us show that for any $x^* \in \mathcal{X}$ the following inequality holds.

$$\limsup_{t\to\infty} \|Ax(t) - y^\delta\| \leq \|Ax^* - y^\delta\|. \tag{A2}$$

Consider for every $t \in [0,\infty)$ the function $e(t) = e(t; x^*) := \frac{1}{2}\|x(t) - x^*\|$. Since $\dot{e}(t) = \langle x(t) - x^*, \dot{x}(t)\rangle$ and $\ddot{e}(t) = \|\dot{x}(t)\|^2 + \langle x(t) - x^*, \ddot{x}(t)\rangle$ for every $t \in [0,\infty)$. Taking into account (4), we get

$$\ddot{e}(t) + \eta\dot{e}(t) + \langle x(t) - x^*, A^*(Ax(t) - y^\delta)\rangle = \|\dot{x}(t)\|^2. \tag{A3}$$

On the other hand, by the convexity inequality of the residual norm square functional $\|Ax(t) - y^\delta\|^2$, we derive

$$\|Ax(t) - y^\delta\|^2 + 2\langle x^* - x(t), A^*(Ax(t) - y^\delta)\rangle \leq \|Ax^* - y^\delta\|^2 \tag{A4}$$

Combine (A3) and (A4) with the definition of $\mathcal{E}(t)$ to obtain

$$\ddot{e}(t) + \eta\dot{e}(t) \leq \frac{1}{2}\|Ax^* - y^\delta\|^2 - \frac{1}{2}\mathcal{E}(t) + \frac{3}{2}\|\dot{x}(t)\|^2.$$

By (A1), $\mathcal{E}(t)$ is nonincreasing, hence, given $t > 0$, for all $\tau \in [0, t]$ we have

$$\ddot{e}(\tau) + \eta\dot{e}(\tau) \leq \frac{1}{2}\|Ax^* - y^\delta\|^2 - \frac{1}{2}\mathcal{E}(t) + \frac{3}{2}\|\dot{x}(\tau)\|^2.$$

By multiplying this inequality with $e^{\eta t}$ and then integrating from 0 to θ, we obtain

$$\dot{e}(\theta) \leq e^{-\eta\theta}\dot{e}(0) + \frac{1-e^{-\eta\theta}}{2\eta}(\|Ax^* - y^\delta\|^2 - \mathcal{E}(t)) + \frac{3}{2}\int_0^\theta e^{-\eta(\theta-\tau)}\|\dot{x}(\tau)\|^2 d\tau.$$

Integrate the above inequality once more from 0 to t together with the fact that $\mathcal{E}(t)$ decreases, to obtain

$$e(t) \leq e(0) + \frac{1-e^{-\eta t}}{\eta}\dot{e}(0) + \frac{\eta t - 1 + e^{-\eta t}}{2\eta^2}(\|Ax^* - y^\delta\|^2 - \mathcal{E}(t)) + h(t), \quad \text{(A5)}$$

where $h(t) := \frac{3}{2}\int_0^t \int_0^\theta e^{-\eta(\theta-\tau)}\|\dot{x}(\tau)\|^2 d\tau d\theta$.

Since $e(t) \geq 0$ and $\mathcal{E}(t) \geq \|Ax(t) - y^\delta\|^2$, it follows from (A5) that

$$\frac{\eta t - 1 + e^{-\eta t}}{2\eta^2}\|Ax(t) - y^\delta\|^2 \leq e(0) + \frac{1-e^{-\eta t}}{\eta}\dot{e}(0) + \frac{\eta t - 1 + e^{-\eta t}}{2\eta^2}\|Ax^* - y^\delta\|^2 + h(t).$$

Dividing the above inequality by $\frac{\eta t - 1 + e^{-\eta t}}{2\eta^2}$ and letting $t \to \infty$, we deduce that

$$\limsup_{t\to\infty}\|Ax(t) - y^\delta\|^2 \leq \|Ax^* - y^\delta\|^2 + \limsup_{t\to\infty}\frac{2\eta}{t}h(t).$$

Hence, for proving (A2), it suffices to show that $h(\cdot) \in L^\infty([0,\infty), \mathcal{X})$. It is obviously held by noting the following inequalities

$$0 \leq h(t) = \frac{3}{2\eta}\int_0^t (1 - e^{-\eta(t-\tau)})\|\dot{x}(\tau)\|^2 d\tau \leq \frac{3}{2\eta}\int_0^\infty \|\dot{x}(\tau)\|^2 d\tau < \infty.$$

From the inequality $\|Ax(t) - y^\delta\| \geq \inf_{x^*\in\mathcal{X}}\|Ax^* - y^\delta\|$, we conclude together with (A2) that

$$\lim_{t\to\infty}\|Ax(t) - y^\delta\| = \inf_{x^*\in\mathcal{X}}\|Ax^* - y^\delta\|. \quad \text{(A6)}$$

Consequently, we have

$$\lim_{t\to\infty}\|Ax(t) - y^\delta\| \leq \|Ax^\dagger - y^\delta\| \leq \delta.$$

Now, let us show the remaining parts of the assertion. Since $\mathcal{E}(t)$ is nonincreasing and bounded from below by $\inf_{x^*\in\mathcal{X}}\|Ax^* - y^\delta\|^2$, it converges as $t \to \infty$. If $\lim_{t\to\infty}\mathcal{E}(t) > \inf_{x^*\in\mathcal{X}}\|Ax^* - y^\delta\|^2$, then $\lim_{t\to\infty}\|\dot{x}(t)\| > 0$ by noting (A6). This contradicts the fact that $x(\cdot) \in L^2([0,\infty), \mathcal{X})$. Therefore, the limit (17) holds and $\dot{x}(t) \to 0$ as $t \to \infty$.

A.2. *Proof of Lemma 2*

(i) Consider for every $t \in [0,\infty)$ the function $e(t) = e(t; x^\dagger) = \frac{1}{2}\|x(t) - x^\dagger\|^2$. Similarly as in (A3), it holds that

$$\ddot{e}(t) + \eta\dot{e}(t) + \langle x(t) - x^\dagger, A^*(Ax(t) - y)\rangle = \|\dot{x}(t)\|^2,$$

which implies that

$$\ddot{e}(t) + \eta\dot{e}(t) + \frac{1}{\|A\|^2}\|A^*(y - Ax(t))\|^2 \leq \|\dot{x}(t)\|^2 \tag{A7}$$

or, equivalently (by using the evolution equation (4)),

$$\ddot{e}(t) + \eta\dot{e}(t) + \frac{\eta}{\|A\|^2}\frac{d\|\dot{x}(t)\|^2}{dt} + \left(\frac{\eta^2}{\|A\|^2} - 1\right)\|\dot{x}(t)\|^2 + \frac{1}{\|A\|^2}\|\ddot{x}\|^2 \leq 0. \tag{A8}$$

By the assumption $\eta \geq \|A\|$, we deduce that

$$\ddot{e}(t) + \eta\dot{e}(t) + \frac{\eta}{\|A\|^2}\frac{d\|\dot{x}(t)\|^2}{dt} \leq 0, \tag{A9}$$

which means that the function $t \mapsto \dot{e}(t) + \eta e(t) + \frac{\eta}{\|A\|^2}\|\dot{x}(t)\|^2$ is monotonically decreasing. Hence, a real number C exists, such that

$$\dot{e}(t) + \eta e(t) + \frac{\eta}{\|A\|^2}\|\dot{x}(t)\|^2 \leq C, \tag{A10}$$

which implies $\dot{e}(t) + \eta e(t) \leq C$. By multiplying this inequality with $e^{\eta t}$ and then integrating from 0 to T, we obtain the inequality

$$e(T) \leq e(0)e^{-\eta T} + C\left(1 - e^{-\eta T}\right)/\eta \leq e(0) + C/\eta.$$

Hence, $e(\cdot)$ is uniform bounded, and, consequently, the trajectory $x(\cdot)$ is uniform bounded.

(ii) follows from Lemma 1.

Now, we prove assertion (iv). Define

$$h(t) = \frac{\eta}{2}\|x(t) - x^\dagger\|^2 + \langle\dot{x}(t), x(t) - x^\dagger\rangle. \tag{A11}$$

By elementary calculations, we derive that

$$\dot{h}(t) = \eta\langle\dot{x}(t), x(t) - x^\dagger\rangle + \langle\ddot{x}(t), x(t) - x^\dagger\rangle + \|\dot{x}(t)\|^2 = \|\dot{x}(t)\|^2 - \|Ax - y\|^2,$$

which implies that (by noting $\dot{\mathcal{E}}(t) = -2\eta\|\dot{x}(t)\|^2$)

$$\dot{\mathcal{E}}(t) + \eta\mathcal{E}(t) + \eta\dot{h}(t) = 0.$$

Integrate the above equation on $[0, T]$ to obtain together, with the nonnegativity of $\mathcal{E}(t)$,

$$\int_0^T \mathcal{E}(t)dt = \frac{1}{\eta}\left(\mathcal{E}(0) - \mathcal{E}(t)\right) - \left(h(t) - h(0)\right) \leq \left(\frac{1}{\eta}\mathcal{E}(0) + h(0)\right) - h(t). \tag{A12}$$

On the other hand, since both $x(t)$ and $x^\dagger$ are uniform bounded, and $\dot{x}(t) \to 0$ as $t \to 0$, a constant M exists such that $|h(t)| \leq M$. Hence, letting $T \to \infty$ in (A12),

we obtain

$$\int_0^\infty \mathcal{E}(t)dt < \infty. \tag{A13}$$

Since $\mathcal{E}(t)$ is non-increasing, we deduce that

$$\int_{T/2}^{T} \mathcal{E}(t)dt \geq \frac{T}{2}\mathcal{E}(T). \tag{A14}$$

Using (A13), the left side of (A14) tends to 0 when $T \to \infty$, which implies that $\lim_{T\to\infty} T\mathcal{E}(T) = 0$. Hence, we conclude $\lim_{T\to\infty} T\|Ax(T) - y\|^2 = 0$, which yields the desired result in (iv).

Finally, let us show the long-term behaviour of $\ddot{x}(\cdot)$. Integrating the inequality (A8) from 0 to T we obtain the fact that there exists a real number C', such that for every $t \in [0, \infty)$

$$\begin{aligned} &\dot{e}(T) + \eta e(T) + \tfrac{\eta}{\|A\|^2}\|\dot{x}(T)\|^2 \\ &\qquad + \left(\tfrac{\eta^2}{\|A\|^2} - 1\right)\int_0^T \|\dot{x}(t)\|^2 dt + \tfrac{1}{\|A\|^2}\int_0^T \|\ddot{x}(t)\|^2 dT \leq C'. \end{aligned} \tag{A15}$$

Since both $e(\cdot)$ and $\dot{e}(\cdot)$ are global bounded (note that $x(\cdot), \dot{x}(\cdot) \in L^\infty([0,\infty), \mathcal{X})$), inequality (A15) gives $\ddot{x}(\cdot) \in L^2([0,\infty), \mathcal{X})$. The relations $\ddot{x}(\cdot) \in L^\infty([0,\infty), \mathcal{X})$ and $\ddot{x}(t) \to 0$ as $t \to \infty$ are obvious by noting assertions (i), (ii), (iv) and the connection equation (4).

Appendix B. Convergence analysis of the second order asymptotical regularization for the case when $\eta \in (0, 2\|A\|]$

B.1. *The underdamped case:* $2\sigma_{j_0+1} < \eta < 2\sigma_{j_0}$

In this case, the solution to the second order differential equation (4) reads

$$x(t) = (1 - A^*Ag^{ab}(t, A^*A))x_0 + \phi^{ab}(t, A^*A)\dot{x}_0 + g^{ab}(t, A^*A)A^*y^\delta,$$

where

$$g^{ab}(t,\lambda) = \begin{cases} g(t,\lambda), & \text{if } \lambda < \eta^2/4, \\ g^b(t,\lambda), & \text{if } \lambda > \eta^2/4, \end{cases} \quad \phi^{ab}(t,\lambda) = \begin{cases} \phi(t,\lambda), & \text{if } \lambda < \eta^2/4, \\ \phi^b(t,\lambda), & \text{if } \lambda > \eta^2/4, \end{cases}$$

where $g(t,\lambda)$ and $\phi(t,\lambda)$ are defined in (24), and

$$\begin{cases} g^b(t,\lambda) = \frac{1}{\lambda}\left\{1 - e^{-\frac{\eta}{2}t}\left[\frac{\eta}{\sqrt{4\lambda-\eta^2}}\sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2}t\right) + \cos\left(\frac{\sqrt{4\lambda-\eta^2}}{2}t\right)\right]\right\}, \\ \phi^b(t,\lambda) = \frac{2}{\sqrt{4\lambda-\eta^2}}e^{-\frac{\eta}{2}t}\sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2}t\right). \end{cases} \tag{B1}$$

As in the overdamped case, we define

$$g_\alpha^{ab}(\lambda) = g^{ab}(1/\alpha, \lambda) \quad \text{and} \quad \phi_\alpha^{ab}(\lambda) = \phi^{ab}(1/\alpha, \lambda). \tag{B2}$$

In this case, the corresponding bias function becomes

$$r_\alpha^{ab}(\lambda) = \begin{cases} r_\alpha(\lambda), & \text{if } \lambda < \eta^2/4, \\ r_\alpha^b(\lambda) := e^{-\frac{\eta}{2\alpha}} \left[\frac{\eta}{\sqrt{4\lambda-\eta^2}} \sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right) + \cos\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right)\right], & \text{if } \lambda > \eta^2/4, \end{cases}$$

where $r_\alpha(\lambda)$ is given in (26).

THEOREM 7 *The functions $\{g_\alpha^{ab}(\lambda), \phi_\alpha^{ab}(\lambda)\}$, defined in (B2), satisfy the conditions $(i)-(iii)$ of Proposition 1.*

Proof. The first requirement in Proposition 1 is obvious. Furthermore, using the inequalities $|\sin\xi| \le |\xi|$ and $e^{-\xi}(\xi+1) \le 1$ we obtain

$$\left| e^{-\frac{\eta}{2\alpha}} \left[\frac{\eta}{\sqrt{4\lambda-\eta^2}} \sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right) + \cos\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right)\right] \right| \le 1, \tag{B3}$$

which implies the second condition in Proposition 1: $|r_\alpha^{ab}(\lambda)| \le \gamma_1^{ab}$ with $\gamma_1^{ab} := \max\left\{ \frac{\eta}{2\sqrt{\eta^2-4\sigma_{j_0+1}^2}} + \frac{1}{2}, 1 \right\}$. Similarly, we have

$$\left|\phi_\alpha^b(\lambda)\right| = \left| \frac{2}{\sqrt{4\lambda-\eta^2}} e^{-\frac{\eta}{2\alpha}} \sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right) \right| \le \alpha^{-1} e^{-\frac{\eta}{2\alpha}} \le \frac{2}{e\eta}, \tag{B4}$$

which means that $|\phi_\alpha^{ab}(\lambda)| \le \gamma_2^{ab}$ with $\gamma_2^{ab} := \max\left\{ \frac{\eta}{2\sqrt{\eta^2-4\sigma_{j_0+1}^2}}, \frac{2}{e\eta} \right\}$.

Now, let us check the third condition in Proposition 1. Using the inequality (B4) we obtain that for $\lambda > \eta^2/4$

$$\frac{1}{\sqrt{\lambda}} \left\{ 1 - e^{-\frac{\eta}{2\alpha}} \left[\frac{\eta}{\sqrt{4\lambda-\eta^2}} \sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right) + \cos\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right)\right] \right\} \le \frac{2}{\sqrt{\lambda}} \le \frac{4}{\eta}.$$

Hence, in the case when $\lambda > \eta^2/4$ and $\alpha \le \eta^2$, we have $\sqrt{\lambda}|g_\alpha^b(\lambda)| \le 4/\eta \le 4/\sqrt{\alpha}$.

Finally, by defining

$$\gamma_*^{ab} = \max\left\{ \sqrt{\eta/(\eta^2 - 4\sigma_{j_0+1}^2)}, 4 \right\} \tag{B5}$$

we can deduce that $\sqrt{\lambda}|g_\alpha^{ab}(\lambda)| \le \gamma_*^{ab}/\sqrt{\alpha}$ for $\alpha \in (0, \bar{\alpha}]$ with $\bar{\alpha} = \eta^2$.

■

PROPOSITION 6 *For all exponents $p > 0$ the monomials $\psi(\lambda) = \lambda^p$ are qualifications with the constants $\gamma_{ab} = \max\{\gamma, \gamma_b\}$ for the second order asymptotical regularization method in the underdamped case, where γ is defined in (29) and*

$$\gamma_b := \frac{\eta + 2\|A\|^2}{2} \left(\frac{2(p+1)}{e\eta} \right)^{p+1} \|A\|^{2p}. \tag{B6}$$

Proof. By Proposition 2, we only need to show that

$$\sup_{\lambda \in (\eta^2/4, \|A\|^2]} |r_\alpha^b(\lambda)|\lambda^p \le \gamma_b \alpha^p \quad \text{and} \quad \sup_{\lambda \in (\eta^2/4, \|A\|^2]} |\phi_\alpha^b(\lambda)|\lambda^p \le \gamma_b \alpha^p \tag{B7}$$

for all $\alpha \in (0, \|A\|^2]$. Set $\xi = \eta/2$ and $p = p' + 1$ in (14) to obtain

$$e^{-\frac{\eta}{2\alpha}} \leq \big(2(p'+1)/(e\eta)\big)^{p'+1} \alpha^{p'+1}. \tag{B8}$$

Then, using (B4) and (B8), we can derive that for all $\alpha \in (0, \|A\|^2]$

$$\begin{aligned}&\left|e^{-\frac{\eta}{2\alpha}}\left[\frac{\eta}{\sqrt{4\lambda-\eta^2}}\sin\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right)+\cos\left(\frac{\sqrt{4\lambda-\eta^2}}{2\alpha}\right)\right]\right|\cdot\lambda^{p'} \leq e^{-\frac{\eta}{2\alpha}}\left(\frac{\eta}{2\alpha}+1\right)\lambda^{p'}\\ &\leq e^{-\frac{\eta}{2\alpha}}\cdot\frac{\eta+2\|A\|^2}{2\alpha}\cdot\|A\|^{2p'} \leq \left(\frac{2(p'+1)}{e\eta}\right)^{p'+1}\cdot\alpha^{p'+1}\cdot\frac{\eta+2\|A\|^2}{2\alpha}\cdot\|A\|^{2p'} = \gamma_b\alpha^{p'},\end{aligned}$$

which yields the first inequality in (B7).

Finally, from the above result, we can deduce that for all $\alpha \in (0, \|A\|^2]$

$$|\phi_\alpha^b(\lambda)|\lambda^{p'} \leq e^{-\frac{\eta}{2\alpha}}\frac{\eta}{2\alpha}\lambda^{p'} \leq e^{-\frac{\eta}{2\alpha}}\left(\frac{\eta}{2\alpha}+1\right)\lambda^{p'} \leq \gamma_b\alpha^{p'},$$

which completes the proof. ■

B.2. *The critical damping case: $\eta = 2\sigma_{j_0}$*

In this case, the solution of (4) is $x(t) = (1 - A^*Ag^{abc}(t, A^*A))x_0 + \phi^{abc}(t, A^*A)\dot{x}_0 + g^{abc}(t, A^*A)A^*y^\delta$, where

$$g^{abc}(t,\lambda) = \begin{cases} g(t,\lambda), & \text{if } \lambda > \eta^2/4,\\ g^b(t,\lambda), & \text{if } \lambda < \eta^2/4,\\ g^c(t,\lambda), & \lambda = \eta^2/4,\end{cases} \quad \phi^{abc}(t,\lambda) = \begin{cases} \phi(t,\lambda), & \text{if } \lambda > \eta^2/4,\\ \phi^b(t,\lambda), & \text{if } \lambda < \eta^2/4,\\ \phi^c(t,\lambda), & \lambda = \eta^2/4,\end{cases}$$

and $g^c(t,\lambda) := \frac{4}{\eta^2}\left\{1 - e^{-\frac{\eta}{2}t}\left(\frac{\eta}{2}t+1\right)\right\}$, $\phi^c(t,\lambda) := te^{-\frac{\eta}{2}t}$.

Define

$$g_\alpha^{abc}(\lambda) = g^{abc}(1/\alpha,\lambda) \quad \text{and} \quad \phi_\alpha^{abc}(\lambda) = \phi^{abc}(1/\alpha,\lambda), \tag{B9}$$

and obtain the corresponding bias function

$$r_\alpha^{abc}(\lambda) = \begin{cases} r_\alpha^a(\lambda), & \text{if } \lambda > \eta^2/4,\\ r_\alpha^b(\lambda), & \text{if } \lambda < \eta^2/4,\\ r_\alpha^c(\lambda) := e^{-\frac{\eta}{2\alpha}}\left(\frac{\eta}{2\alpha}+1\right), & \text{if } \lambda = \eta^2/4.\end{cases}$$

THEOREM 8 *The functions $\{g_\alpha^{abc}(\lambda), \phi_\alpha^{abc}(\lambda)\}$, given in (B9), satisfy the conditions $(i) - (iii)$ of Proposition 1.*

Proof. By Theorem 7, we only need to check the case when $\lambda = \eta^2/4$. In this case, it is easy to verify that $\lim_{\alpha\to0}\phi_\alpha(\lambda) = \lim_{\alpha\to0} e^{-\frac{\eta}{2\alpha}}/\alpha = 0$, $\lim_{\alpha\to0} r_\alpha(\lambda) = \lim_{\alpha\to0} e^{-\frac{\eta}{2\alpha}}\left(\frac{\eta}{2\alpha}+1\right) = 0$ and $|\phi_\alpha(\lambda)| \leq \frac{2}{e\eta}$ and $|r_\alpha(\lambda)| \leq 1$ for all $\alpha > 0$. Finally, by the inequality (assume that $\alpha \leq \eta^2$) $\frac{1}{\sqrt{\lambda}}\left\{1 - e^{-\frac{\eta}{2\alpha}}\left(\frac{\eta}{2\alpha}+1\right)\right\} \leq \frac{2}{\sqrt{\lambda}} = \frac{4}{\eta} \leq \frac{4}{\sqrt{\alpha}}$, and Theorem 7, we complete the proof with $\sqrt{\lambda}|g_\alpha(\lambda)| \leq \gamma_*^{ab}/\sqrt{\alpha}$ for $\alpha \in (0, \bar{\alpha}]$, $\bar{\alpha} = \eta^2$, and γ_*^{ab} is defined in (B5). ■

PROPOSITION 7 *For all exponents $p > 0$ the monomials $\psi(\lambda) = \lambda^p$ are qualifications with the constants $\gamma_{abc} = \max\{\gamma, \gamma_b, \gamma_c\}$ for the second order asymptotical*

regularization method in the critical damping case, where

$$\gamma_c := \frac{\eta + 2\|A\|^2}{2} \left(\frac{p+1}{e}\right)^{p+1} \left(\frac{\eta}{2}\right)^{p-2} \max\left(\frac{\eta}{2}, 1\right), \tag{B10}$$

and the constants γ and γ_b are defined in (29) and (B6) respectively.

Proof. By Propositions 2 and 6, we only need to show that for all $\alpha \in (0, \|A\|^2]$

$$|r_\alpha(\eta^2/4)|(\eta^2/4)^p \leq \gamma_c \alpha^p \text{ and } |\phi_\alpha(\eta^2/4)|(\eta^2/4)^p \leq \gamma_c \alpha^p.$$

By (B9) and elementary calculations, we derive that

$$\begin{aligned}
&\left|r_\alpha\left(\tfrac{\eta^2}{4}\right)\right| \left(\tfrac{\eta^2}{4}\right)^p = e^{-\frac{\eta}{2\alpha}} \left(\tfrac{\eta}{2\alpha} + 1\right) \left(\tfrac{\eta^2}{4}\right)^p \leq \left(\left(\tfrac{2(p+1)}{e\eta}\right)^{p+1} \alpha^{p+1}\right) \tfrac{\eta+2\|A\|^2}{2\alpha} \left(\tfrac{\eta^2}{4}\right)^p \\
&= \left\{\tfrac{\eta+2\|A\|^2}{2} \left(\tfrac{2(p+1)}{e\eta}\right)^{p+1} \left(\tfrac{\eta^2}{4}\right)^p\right\} \alpha^p = \left\{\tfrac{\eta+2\|A\|^2}{2} \left(\tfrac{p+1}{e}\right)^{p+1} \left(\tfrac{\eta}{2}\right)^{p-1}\right\} \alpha^p \leq \gamma_c \alpha^p,
\end{aligned}$$

and

$$\begin{aligned}
&\left|\phi_\alpha\left(\tfrac{\eta^2}{4}\right)\right| \left(\tfrac{\eta^2}{4}\right)^p = e^{-\frac{\eta}{2\alpha}} \tfrac{\eta}{2\alpha} \tfrac{2}{\eta} \left(\tfrac{\eta^2}{4}\right)^p \leq e^{-\frac{\eta}{2\alpha}} \left(\tfrac{\eta}{2\alpha} + 1\right) \left(\tfrac{\eta}{2}\right)^{2p-1} \\
&\leq \left(\left(\tfrac{p+1}{e}\right)^{p+1} \alpha^{p+1}\right) \tfrac{\eta+2\|A\|^2}{2\alpha} \left(\tfrac{\eta}{2}\right)^{2p-1} \leq \gamma_c \alpha^p,
\end{aligned}$$

which yields the required result. ■