Strong Solvability of a Variational Data Assimilation Problem for the Primitive Equations of Large-Scale Atmosphere and Ocean Dynamics

Peter Korn[1]

Abstract

For the primitive equations of large-scale atmosphere and ocean dynamics, we study the problem of determining by means of a variational data assimilation algorithm initial conditions that generate strong solutions which minimize the distance to a given set of time-distributed observations. We suggest a modification of the adjoint algorithm whose novel elements is to use norms in the variational cost functional that reflects the H^1-regularity of strong solutions of the primitive equations. For such a cost functional, we prove the existence of minima and a first-order adjoint condition for strong solutions that provides the basis for computing these minima. We prove the local convergence of a gradient-based descent algorithm to optimal initial conditions using the second-order adjoint primitive equations. The algorithmic modifications due to the H^1-norms are straightforwardly to implement into a variational algorithm that employs the standard L^2-metrics.

1 Introduction, Statement of Problem and Overview of Results

Data assimilation is a computational technique that aims at blending a dynamical model of a physical process with observational data of this process while at the same time preserving the integrity of the model and making optimal use of the observa-

[1] Max-Planck-Institute for Meteorology, Hamburg, Germany

tional information. It constitutes a fundamental technique for modeling real-world phenomena. Numerical weather prediction and ocean state estimation are examples of scientific disciplines that rely on data assimilation (Kalnay 2003; Wunsch 1997; Forget et al. 2015). Data assimilation determines a best estimate of the unknown current state of the atmosphere or the ocean by steering the model trajectory toward time distributed observations of the past. The estimate of the current state is used as initial condition for a prediction. The rationale behind is, the better the estimate of the present state, the better the prediction of the future evolution.

The data assimilation algorithm that we consider here is known as *"4D-var"* or *"adjoint method"* of optimal control and belongs to the class of variational algorithms. Variational data assimilation algorithms strive for minimizing the distance between observations and model solution by solving an optimization problem for an appropriately chosen cost functional. The initial condition acts as a control variable of the optimization problem that fits the model trajectory over a finite time horizon to the observations. In contrast, feedback control algorithms such as *"continuous data assimilation " or "nudging"* drive the model over an infinite time horizon toward observations by means of a feedback term (Kalnay 2003). Recently, these algorithms have been improved and rigorous results have been proven for a range of fluid dynamical equations, including the Navier–Stokes or the planetary geostrophic equations (see, e.g., Azouani and Titi 2014; Foias et al. 2016; Desamsetti et al. 2019; Farhat et al. 2016; Korn 2009) and the primitive equation (Pei 2019). A complementary class of assimilation algorithms is given by stochastic data assimilation such as the Kalman filter and its variants (Evensen 2003; Reich and Cotter 2015; Houtekamer and Zhang 2016), for mathematical aspects we refer to Schillings and Stuart (2017, 2018). Variational as well as stochastic algorithms are of great relevance in atmosphere-ocean science. Research in this area focuses on improvement and extension of computational algorithms in order to improve numerical weather predictions or ocean state estimation. The complexity of data assimilation algorithms and their applications have reached an impressive level (see, e.g., Bonavita et al. 2018; Isaksen et al. 2010; Gauthier et al. 2007). It is therefore important to understand fundamental properties of these algorithms. Results that are similar in spirit to our work can be found in Agoshkov and Ipatova (2007), where a variational algorithm was analyzed that uses observations of sea surface temperature and of sea surface elevation to determining the control variables of heat flux and water flux that occur in the surface boundary conditions of the oceanic primitive equations. The authors used a cost functional based on the L^2-norm and showed the existence of minimizers. In Shutyaev (1997), the more general setting of a data assimilation problem for quasi-linear evolution equations in Hilbert spaces was studied. To illustrate the general theory the existence of minimizers and the convergence of an iterative adjoint algorithm was established for a two-dimensional barotropic fluid. For fluid dynamical equations such as the 2D Navier–Stokes or the 3D Navier–Stokes-α equations similar results on optimal control using initial conditions, boundary conditions or external forcing as control variable can be found in Fursikov (1999), Abergel and Temam (1990) and Korn (2009). In Korn (2019), an idealized coupled atmosphere-ocean model was introduced and based on the regularity of the coupled equations the existence of optimal initial conditions and the convergence of a steepest descent method was proven.

In the 4D-var algorithm, the crucial modeling decision is how the distance to the observational data and to the background state is measured, i.e., how the cost functional is specified. The classical choice is the L^2-norm. In this paper, we suggest to use for the primitive equations the $\mathcal{H}^1$-norm. This norm in the cost functional reflects the $\mathcal{H}^1$-regularity of strong solutions of the primitive equations. The optimization determines initial conditions for strong solutions of the primitive equations. With such a choice, we take advantage of a priori knowledge about the dynamics of the primitive equations.

The use of alternative metrics beyond $L^2([0, T], \mathcal{L}^2(\Omega))$ is not new in optimal control of fluids [for a review see Medjo et al. (2008)]. In Bewley et al. (2001), for example, enstrophy-based metrics and Sobolev-type cost functionals were discussed in the context of turbulence control via boundary forcing for a channel flow. The purpose here is to obtain additional regularizing effect due to derivatives in the cost functional. In contrast, in our work here, the goal is not to introduce a regularization of the data assimilation problem but to demonstrate that optimal initial conditions for strong solutions of the primitive equations assimilation problem necessitates the use of $\mathcal{H}^1$-norms. We refer to this property as *strong solvability of the data assimilation problem*. The new and fundamental but decisive element in the suggested formulation of the data assimilation algorithm is to use in the cost functional metrics that are tailored to the regularity of the primitive equations. This paper describes for the primitive equations the consequences of such a choice.

1.1 The Primitive Equations

The primitive equations are a central set of model equations for atmosphere and ocean dynamics. These equations approximate the large-scale dynamics of ocean and atmosphere and describe ocean or atmosphere as thin layer of a rotating, incompressible fluid, together with the Boussinesq approximation and the hydrostatic approximation. The primitive equations are derived as an approximation of the rotating Navier–Stokes equations Li and Titi (1992). The state vector of the primitive equations consists of a horizontal velocity field $\mathbf{v} = (u, v)$ and a tracer such as temperature θ. More specifically, denote by $\Omega := M \times (-h, 0)$ a cylindrical domain, where $h > 0$ denotes the depth of the ocean and $M \subseteq \mathbb{R}^2$ is bounded, with a smooth boundary ∂M. The primitive equations are given by

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} + w\partial_z \mathbf{v} + \nabla p + f\vec{k} \times \mathbf{v} - \frac{1}{Re_1}\triangle \mathbf{v} - \frac{1}{Re_2}\partial_{zz}^2 \mathbf{v} = F_{\mathbf{v}}, \qquad (1a)$$

$$\partial_z p + g\rho = 0, \qquad (1b)$$

$$\nabla \cdot \mathbf{v} + \partial_z w = 0, \qquad (1c)$$

$$\partial_t \theta + (\mathbf{v} \cdot \nabla)\theta + w\partial_z \theta - \frac{1}{Rt_1}\triangle \theta - \frac{1}{Rt_2}\partial_{zz}^2 \theta = F_\theta, \qquad (1d)$$

$$\rho = \rho(\theta) := \rho_0 - \alpha(\theta - \theta_0), \qquad (1e)$$

where w denotes the vertical velocity, f the Coriolis parameter, ρ denotes an equation of state with a mean density $\rho_0 > 0$, a mean temperature $\theta_0 > 0$ and a coefficient $\alpha > 0$

of thermal expansion, $F_{\mathbf{v}}$, F_θ are forcing terms. The numbers Re_1, $Re_2 > 0$ denote the horizontal and vertical Reynolds number that represent the viscosity coefficients. The numbers Rt_1, $Rt_2 > 0$ are the horizontal and vertical mixing coefficients for temperature. The operators ∇, $\nabla\cdot$ and $\triangle$ denote the horizontal gradient, divergence and Laplacian, respectively. The partial derivative with respect to the vertical direction is denoted by ∂_z.

The boundary $\partial\Omega$ consists of a lateral boundary Γ_s, a bottom boundary Γ_b and a surface boundary Γ_u, i.e., $\partial\Omega = \Gamma_s \cup \Gamma_b \cup \Gamma_u$. The corresponding boundary conditions are

$$on\ \Gamma_b := \{(x, y, z) \in \bar{\Omega} : z = -h\} : \mathbf{v} = 0, w = 0, \tag{2}$$

$$[\nabla_3\theta] \cdot n_3 = 0, \tag{3}$$

$$on\ \Gamma_s := \{(x, y, z) \in \bar{\Omega} : (x, y) \in \partial M\} : \mathbf{v} = 0, w = 0, \tag{4}$$

$$[\nabla_3\theta] \cdot \vec{n} = 0, \tag{5}$$

$$on\ \Gamma_u := \{(x, y, z) \in \bar{\Omega} : z = 0\} : \partial_z\mathbf{v} = h\tau, w = 0, \tag{6}$$

$$[\nabla_3\theta] \cdot n_3 = -k_\theta(\theta - \theta^*), \tag{7}$$

where τ is a given 2D wind stress field, $\vec{n}$ the normal vector at Γ_s, n_3 the vertical Cartesian unit vector, and θ^* is a given temperature field at the sea surface. The velocity $\mathbf{v}$ and the potential temperature θ can be modified at Γ_s by adding a τ- and θ^*-dependent contribution such that we can assume without loss of generality homogeneous boundary conditions at the ocean surface (see Lions et al. 1992, Section 2.4., Cao and Titi 2007 p. 248, Cao and Titi 2003, p. 202)

$$\partial_z\mathbf{v} = 0 \ \text{ and } \ \partial_z\theta = 0, \quad \text{on } \Gamma_u. \tag{8}$$

The mathematical analysis of primitive equations dates back to Lions et al. (1992), where the existence of *global in time weak solutions* was proven for square integrable initial conditions. Weak solutions satisfy the regularity properties $(\mathbf{v}, \theta) \in L^\infty([0, T], \mathcal{L}^2(\Omega)) \cap L^2([0, T], \mathcal{H}^1(\Omega))$[1]. The uniqueness of weak solutions remains still an open question (for weak solutions see also remark 1 in Section 2). The local existence of strong solutions was shown by Guillén-González et al. (2001). In a seminal paper, Cao and Titi (2007) have proven the global existence and the uniqueness of *strong solutions* for arbitrary initial data in the Sobolev space $\mathcal{H}^1$. Strong solutions possess more regularity than weak solutions, namely $(\mathbf{v}, \theta) \in L^\infty([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega))$. The notion of *"strong solution"* is introduced in below Definition 2.1. The well-posedness result Cao and Titi (2007) uses Neumann boundary conditions for velocity on top and bottom of the domain, and a slip boundary condition at lateral sides. A different proof was given by Kobelkov (2007). This breakthrough has been followed by well-posedness results in H^1 for

---

[1] We use the notation $H^1(\Omega)$, $\mathbf{H}^1(\Omega)$ and $\mathcal{H}^1(\Omega)$ to distinguish between the Sobolev space with square-integrable derivatives for temperature, for velocity and for the product space of velocity and temperature. The same notation is used for the Sobolev space with square-integrable second derivatives H^2, $\mathbf{H}^2$, $\mathcal{H}^2$ and for spaces of square integrable functions, L^2, $\mathbf{L}^2$, $\mathcal{L}^2$. The definitions are given in Sect. 2.

Neumann boundary conditions at the surface and Dirichlet boundary conditions at the bottom and at the lateral boundary by Kukavica and Ziane (2007). The existence of an attractor of the primitive equations was shown in Chueshov (2014) and Ju (2007). The well-posedness of the primitive equations for L^p-spaces is proven in Hieber and Kashiwabara (2016). In this work, we use Dirichlet boundary conditions at lateral sides and bottom of the domain and Neumann boundary conditions at the surface, because these are the most common boundary conditions in ocean general circulation models but our results translate to the other boundary conditions mentioned above.

The $\mathcal{H}^1$-regularity forms the mathematical basis for the specification of the data assimilation problem. We state now the central theorem for later reference. The space $\mathcal{V}$ occurring in the Theorem consists of pairs $(\mathbf{v}, \theta)$ of velocity and temperature in $\mathcal{H}^1(\Omega)$ that satisfy the boundary conditions (2)–(7) and is defined in Sect. 2 [see (16)].

Theorem 1.1 (Global Well-Posedness (Cao and Titi 2007; Kukavica and Ziane 2007)) *Let a time interval $[0, T]$, with $T > 0$, be given. Let the forcing satisfy $F_{\mathbf{v}}, F_{\theta} \in L^2([0, T], \mathcal{L}^2(\Omega))$. If the initial conditions for velocity and temperature satisfy $(\mathbf{v}_0, \theta_0) \in \mathcal{V}$, then there exists a unique strong solution $(\mathbf{v}, \theta) \in C([0, T], \mathcal{H}^1(\Omega)) \cap L^2[0, T], \mathcal{H}^2(\Omega))$ of the system (1) on the interval $[0, T]$ that depends continuously on the initial data.*

The regularity of solutions of the primitive equations, described in Theorem 1.1, still leaves room for a broad spectrum of spatiotemporal scales that comprises local and sudden events with large gradients such as fronts. For an analysis of the admissible scales in terms of Nusselt, Rayleigh and Reynolds numbers, we refer to Gibbon and Holm (2011, 2013).

1.2 The Data Assimilation Problem

We now proceed by describing the data assimilation problem for the primitive equations. Let $\mathcal{X} = L^{\infty}([0, T], \mathcal{H}^1(\Omega))$ be the "state space" of strong solutions of the primitive equations with initial conditions $\mathcal{X}_0 = \mathcal{H}^1(\Omega)$. Let time distributed observations ψ_{obs} are given. In general, the observations reside in an "observational space" $\mathcal{Y}$ and an observation operator H maps $\psi_{\mathrm{obs}} \in \mathcal{Y}$ in the model space $\mathcal{X}$. In this work, we assume that the observations are already mapped to the model space, i.e., $\psi_{\mathrm{obs}} \in \mathcal{X}$.

The *data assimilation problem* consists in determining, for given observations $\psi_{\mathrm{obs}} \in \mathcal{X}$ an initial condition $\bar{\psi}_0 \in \mathcal{X}_0$ such that

$$\mathcal{J}(\bar{\psi}_0, \psi_{\mathrm{obs}}) = \min_{\psi_0 \in \mathcal{X}_0} \mathcal{J}(\psi_0, \psi_{\mathrm{obs}})$$

and

the trajectory $\psi(\bar{\psi}_0) \in \mathcal{X}$ *satisfies the primitive equations* (1)

with initial condition $\bar{\psi}_0$.

(9)

The cost functional $\mathcal{J}$ in (9) is defined as a sum of a background and an observational term

$$\mathcal{J}(\psi_0, \psi_{\text{obs}}) := \mathcal{J}_b(\psi_0) + \mathcal{J}_{\text{obs}}(\psi_0, \psi_{\text{obs}}). \tag{10}$$

The observational part $\mathcal{J}_{\text{obs}}$ of the cost functional measure the distance between the model state and the observations and is defined by

$$\mathcal{J}_{\text{obs}}(\psi_0, \psi_{\text{obs}}) := \int_T \langle \mathcal{R}(\mathcal{M}[\psi_0] - \psi_{\text{obs}}), \mathcal{M}[\psi_0] - \psi_{\text{obs}} \rangle_{\mathcal{X}} \mathrm{d}t. \tag{11}$$

Here, $\mathcal{M}$ denotes the *model operator* that advances an initial condition ψ_0 to the model state $\psi(t)$ at time t, i.e., $\psi(t) = \mathcal{M}[\psi_0](t)$. The observational error covariance operator $\mathcal{R}$ provides a statistical weighting of the model-observation misfit, according to the quality of the observations. The scalar product in (11) acts on the spatial variable. The background term $\mathcal{J}_b$ of the cost functional in (10) is defined by

$$\mathcal{J}_b(\psi_0) := \mathcal{J}_b(\psi_0, \psi_{\text{back}}) := \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \psi_0 - \psi_{\text{back}} \rangle_{\mathcal{X}_0}, \tag{12}$$

where $\psi_{\text{back}} \in \mathcal{X}_0$ is a given background state such as the outcome of a previous forecast and $\mathcal{B}$ denotes the model error covariance operator. The background term $\mathcal{J}_b$ incorporates prior information about the system but it can also be interpreted as a Tikhonov regularization of $\mathcal{J}$. The model error covariance operator $\mathcal{B}$ provides a weighting according to the estimated model error.

As a consequence of the nonlinearity of the primitive equations, encoded in the model operator $\mathcal{M}$ in (11), we can for the data assimilation problem (9) not expect a global minimum but at most multitude of local minima. In principle, the uniqueness of the optimization can be enforced by weighting the background and observational parts of the cost functional appropriately but this may come at the expense of discarding the observational information [see Theorem 7 in Korn (2009) for a result in this respect].

1.3 Overview of Results

The relevance of the primitive equations for atmosphere/ocean science and climate research together with the fact that these equations constitute one of the rare examples of a 3D fluid dynamical equation with a global well-posedness result poses the question how this knowledge can be incorporated into a 4D-var data assimilation algorithm. We advocate a formulation of the variational cost functional (10) that reflects the regularity of the underlying dynamical equations. This philosophy has already been demonstrated in the context of an idealized coupled atmosphere-ocean model (Korn 2019).

The classical formulation of 4D-Var based on the $\mathcal{L}^2$-norm in space and with initial conditions in $\mathcal{L}^2(\Omega)$ is associated with the notion of weak solutions. These solutions describe the dynamics in the L^2-space. For weak solutions the uniqueness as well as the continuous dependency on the initial conditions is unknown but for the adjoint approach of variational data assimilation not only continuity but differentiability with

respect to the initial condition is required. The situation changes completely if we employ the notion of *strong solutions* of the primitive equations, for which uniqueness and continuous dependency of the solution on the initial condition are established.

In view of the data assimilation problem strong solutions require optimal initial conditions with respect to the observations in the space $\mathcal{H}^1(\Omega)$. These initial conditions generates dynamics in $L^\infty([0, T], \mathcal{H}^1) \cap L^2([0, T], \mathcal{H}^2(\Omega))$ and consequently we measure the distance to observations in this space, i.e., we use in the background term $\mathcal{J}_b$ of the cost functional (10) the $\mathcal{H}^1$-norm and in the observational term $\mathcal{J}_{obs}$ the $L^2([0, T], \mathcal{H}^1(\Omega))$-norm. This imposes stronger constraints through the higher derivatives than the L^2-norm. The $L^2([0, T], \mathcal{H}^2(\Omega))$-norm can *not* be used in the observational term of the cost functional, because the degree of the spatial derivative in this term is linked via the Gateaux-derivative of the cost functional to a differential operator that is part of the forcing of the adjoint equations. The second-order derivative in the spatial H^2-norm would result in an adjoint forcing that is no longer square-integrable. For details, we refer to Theorem 4.4 and remark 2 in Sect. 4.2.

In Lemma 3.2, we prove that strong solutions are differentiable with respect to the initial conditions. For strong solutions of the primitive equations, we prove in Theorem 4.1 that local minimizers of the cost functional exists. Theorem 4.4 gives a first-order necessary adjoint condition for these minimizers that provides the basis for their computation by means of an adjoint based optimizations. For strong solutions, we prove the convergence of a steepest-descent algorithm in Theorem 5.3, provided one uses a suitable starting point for the iterative process. Pure Newton methods that allow global convergence have never been used for realistic applications in numerical weather prediction or ocean state estimation due to their prohibitive computational costs. Variants such as quasi-Newton methods are not considered in this paper.

Our formulation of the variational data assimilation for the primitive equations does only modify the optimization framework. Affected are the forcing of the adjoint equation and the gradient that results from an integration of the adjoint equation, while the dynamics of the primitive equations remain untouched. The overall algorithmic structure of an existing 4D-Var code is also left intact. The choice of the $\mathcal{H}^1$-norm in the background term results for a computational gradient-based algorithm in a filter of the gradient through an inverse Helmholz operator. The $L^2([0, T], \mathcal{H}^1(\Omega))$-norm in the observational term implies a modification of the forcing of the adjoint equation with a Helmholtz operator applied to the model-data-misfit. Both of these changes are easy to integrate into an existing 4D-Var algorithm (see Fig.1).

The impact of regularity-tailored cost functionals on the quality as well as on the performance of the 4D-Var algorithm has to be assessed in numerical simulations. In particular the potential of the method for controlling turbulent flows has to be investigated. This comprises the comparison of the quality of the $\mathcal{H}^1$- with the $\mathcal{L}^2$-4D-Var data assimilation and the potential to resolve issues of ($\mathcal{L}^2$-) 4D-Var such as strong adjoint sensitivities or large number of local minima (see, e.g., Hoteit et al. 2005, (Köhl and Willebrand 2002; Lorenc 2003). The experimental investigation of these topics is beyond the scope of this paper.

2 Functional Analytic Setup

Let us recall the following basic notation

- $\vec{v} := (\mathbf{v}, w)$ velocity field with horizontal velocity $\mathbf{v} := (u, v)$ and vertical velocity w
- θ temperature
- $\nabla := (\partial_x, \partial_y), \ \nabla_3 := (\partial_x, \partial_y, \partial_z)$
- $\nabla \cdot \mathbf{v} := \partial_x u + \partial_y v, \ \nabla_3 \cdot \mathbf{v} = \partial_x u + \partial_y v + \partial_z w$

$$(13)$$

We define the following fundamental function spaces for velocity and temperature

$$
\begin{aligned}
\tilde{V}_{\mathbf{v}} :&= \left\{ \mathbf{v} \in (\mathbf{C}^\infty(\bar{\Omega}))^2 : \ \mathbf{v} \text{ satisfies boundary condition (2), (4), (6)} \right. \\
&\left. \quad \text{and } \nabla \cdot \mathbf{v} + \partial_z w = 0 \right\}, \\
\tilde{V}_\theta :&= \left\{ \theta \in C^\infty(\bar{\Omega}) : \theta \text{ satisfies boundary condition (3), (5), (7)} \right\}.
\end{aligned}
$$

$$(14)$$

Denote by $V_{\mathbf{v}}, V_\theta$ the closure of $\tilde{V}_{\mathbf{v}}$ in the Sobolev space $\mathbf{H}^1(\Omega)$ and of $\tilde{V}_\theta$ in $H^1(\Omega)$. The norm of $V_{\mathbf{v}}$ and V_θ is

$$
\begin{aligned}
||\mathbf{v}||_{\mathbf{H}^1}^2 :&= \int_\Omega |\mathbf{v}(x, y, z)|^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega |\nabla_3 \mathbf{v}(x, y, z)|^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z, \\
\text{and} \ \ ||\theta||_{H^1}^2 :&= \int_\Omega |\theta(x, y, z)|^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega |\nabla_3 \theta(x, y, z)|^2 \mathrm{d}x\mathrm{d}y\mathrm{d}z.
\end{aligned}
$$

$$(15)$$

The product space by

$$\mathcal{V} := V_{\mathbf{v}} \times V_\theta. \tag{16}$$

is equipped with the sum of the norms $|| \cdot ||_{\mathcal{H}^1} := || \cdot ||_{\mathbf{H}^1} + || \cdot ||_{H^1}$. We denote the L^2-norm in the velocity and temperature space by $|| \cdot ||_{\mathbf{L}^2}$ and $|| \cdot ||_{L^2}$, and in the velocity-temperature product space by $|| \cdot ||_{\mathcal{L}^2} := || \cdot ||_{\mathbf{L}^2} + || \cdot ||_{L^2}$. The H^2-norm in the velocity-temperature product space is denoted by $|| \cdot ||_{\mathcal{H}^2} := || \cdot ||_{\mathbf{H}^2} + || \cdot ||_{H^2}$.

In (1), the vertical velocity w is determined by the constraint (1c) and can by virtue of the boundary conditions be expressed as

$$w(x, y, z, t) = -\int_{-z}^0 \nabla \cdot \mathbf{v}(x, y, \xi, t) \, \mathrm{d}\xi = -\nabla \cdot \int_{-z}^0 \mathbf{v}(x, y, \xi, t) \, \mathrm{d}\xi. \tag{17}$$

The pressure term can with the hydrostatic balance be reformulated as

$$p(x, y, z, t) = \int_{-h}^z \theta(x, y, \xi, t) \, \mathrm{d}\xi + p_s(x, y, t), \tag{18}$$

where p_s is the surface pressure and has to be determined.

Definition 2.1 (Strong Solution) (*i*) Let initial conditions $(\mathbf{v}_0, \theta_0) \in \mathcal{V}$ be given, and let $[0, T]$, $T > 0$, be a time interval. The vector $(\mathbf{v}, \theta)$ is called a strong solution of (1) on $[0, T]$ with boundary conditions (2)–(8) if

$$\mathbf{v} \in C([0, T], V_{\mathbf{v}}) \cap L^2([0, T], \mathbf{H}^2(\Omega)),$$
$$\theta \in C([0, T], V_\theta) \cap L^2([0, T], H^2(\Omega)),$$
$$\partial_t \mathbf{v} \in L^2([0, T], \mathbf{L}^2(\Omega)),$$
$$\partial_t \theta \in L^2([0, T], L^2(\Omega)),$$

and if it satisfies in the sense of distributions for all $\Phi \in (C^\infty(\bar{\Omega}))^2$ and $\phi \in C^\infty(\bar{\Omega})$

$$\int_\Omega \partial_t \mathbf{v} \cdot \Phi \, dx + \int_\Omega [(\mathbf{v} \cdot \nabla)\mathbf{v}] \cdot \Phi \, dx + \int_\Omega f \vec{k} \times \mathbf{v} \Phi \, dx dy dz$$
$$- \int_\Omega \left[\left(\nabla \cdot \int_{-h}^z \mathbf{v}(x, y, \xi, t) \, d\xi \right) \partial_z \mathbf{v} \right] \cdot \Phi \, dx dy dz$$
$$- \int_\Omega \left[\nabla \int_{-h}^z g\rho(x, y, \xi, t) \, d\xi \right] \cdot \Phi \, dx dy dz + \int_{\Gamma_u} (\nabla p_s) \cdot \Phi \, dx dy dz$$
$$+ \int_\Omega \frac{1}{Re_1} \nabla \mathbf{v} \cdot \nabla \Phi + \frac{1}{Re_2} \partial_z \mathbf{v} \cdot \partial_z \Phi \, dx dy dz = F_{\mathbf{v}}, \tag{19a}$$
$$\int_\Omega \partial_t \theta \phi \, dx dy dz + \int_\Omega (\mathbf{v} \cdot \nabla)\theta \phi \, dx dy dz$$
$$- \int_\Omega \left[\nabla \cdot \int_{-h}^0 \mathbf{v}(x, y, \xi, t) \, d\xi \right] \partial_z \theta \, \phi \, dx dy dz$$
$$+ \frac{1}{Rt_1} \int_\Omega \nabla_3 \theta \cdot \nabla_3 \phi \, dx dy dz + \frac{1}{Rt_2} \int_\Omega \partial_z \theta \partial_z \phi \, dx dy dz = F_\theta,$$
with $\rho := \rho(\theta) := \rho_0 - \alpha(\theta - \theta_0)$, $\tag{19b}$

where w denotes the vertical velocity, f the Coriolis parameter and $\rho_0, \theta_0 > 0$ are a mean density and a mean temperature, respectively, and $\alpha > 0$ the thermal expansion coefficient, $F_{\mathbf{v}}, F_\theta$ denote the external forcing. The numbers $Re_1, Re_2 > 0$ denote the horizontal and vertical Reynolds number that represent the viscosity coefficients. The numbers $Rt_1, Rt_2 > 0$ are the horizontal and vertical mixing coefficients for temperature.
(*ii*) We define the *state vector* ψ of the primitive equation as pair of velocity and temperature variables

$$\psi := (\mathbf{v}, \theta), \tag{20}$$

where $(\mathbf{v}, \theta)$ denote a strong solution of the primitive equation. Furthermore, we define the nonlinear terms as follows:

$$\mathcal{B}(\psi, \psi) := (\mathcal{B}_v(\mathbf{v}, \mathbf{v}), \mathcal{B}_\theta(\theta, \theta)) \tag{21}$$

$$\text{with} \quad \mathcal{B}_v(\mathbf{v}, \mathbf{v}) := (\mathbf{v} \cdot \nabla)\mathbf{v} - \left(\nabla \cdot \int_{-h}^{z} \mathbf{v}(x, y, \xi, t) \, d\xi \right) \partial_z \mathbf{v}, \qquad (22)$$

$$\mathcal{B}_\theta(\theta, \theta) := (\mathbf{v} \cdot \nabla)\theta - \left(\nabla \cdot \int_{-h}^{z} \mathbf{v}(x, y, \xi, t) \, d\xi \right) \partial_z \theta. \qquad (23)$$

The dissipation operators are denoted by

$$D\psi := (D_v \mathbf{v}, D_\theta \theta), \qquad (24)$$

$$\text{with} \quad D_v \mathbf{v} := -\frac{1}{Re_1} \triangle \mathbf{v} - \frac{1}{Re_2} \partial_{zz}^2 \mathbf{v}, \qquad (25)$$

$$D_\theta := -\frac{1}{Rt_1} \triangle \theta - \frac{1}{Rt_2} \partial_{zz}^2 \theta. \qquad (26)$$

The linear terms are given by

$$L\psi := (L_v \mathbf{v}, 0) \qquad (27)$$

$$\text{with} \quad L_v \mathbf{v} := f\vec{k} \times \mathbf{v} + \int_{-h}^{z} \theta(x, y, \xi, t) \, d\xi) + \nabla p_s. \qquad (28)$$

With this notation, we write the primitive equations (1) in the form

$$\partial_t \psi + \mathcal{B}(\psi, \psi) + L\psi + D\psi = 0, \qquad (29)$$

and interpret this equation in the distributional sense.

Remark 1 (Weak Solutions) Weak solutions of the primitive equations are defined with less regularity requirements. They satisfy (19) in the sense of distributions but with $(\mathbf{v}, \theta) \in L^\infty([0, T], \mathcal{L}^2(\Omega)) \cap L^2([0, T], \mathcal{H}^1(\Omega))$. The global existence of weak solutions with initial data in $(\mathbf{v}_0, \Theta_0) \in \mathcal{L}^2(\Omega)$ has been shown in Lions et al. (1992). The uniqueness is still open. Uniqueness of weak solutions with initial conditions in the space of continuous functions was proven in Kukavica et al. (2014).

We recall a set of inequalities that are used in the sequel. The Ladyshenzkaya inequalities in $\mathbb{R}^2$ are valid for $\phi \in H^1(M)$ and for a non-dimensional constant C_0

$$||\phi||_{L^4(M)} \leq C_0 ||\phi||_{L^2(M)}^{1/2} ||\phi||_{H^1(M)}^{1/2}. \qquad (30)$$

In $\mathbb{R}^3$, it holds for $u \in H^1(\Omega)$

$$||u||_{L^3(\Omega)} \leq C_0 ||u||_{L^2(\Omega)}^{1/2} ||u||_{H^1(\Omega)}^{1/2},$$
$$||u||_{L^6(\Omega)} \leq C_0 ||u||_{H^1(\Omega)}. \qquad (31)$$

Lemma 2.2 (Cao and Titi 2003) *Let $u = (u_1, u_2) \in H^2(\Omega)$ and $f, g \in H^1(\Omega)$. Then*

$$\left| \int_\Omega \left(\nabla \cdot \int_{-h}^z u(x, y, \xi, t) \, d\xi \right) fg \, dx\,dy\,dz \right|$$
$$\leq C \|f\|_{L^2} \|u\|_{H^1}^{1/2} \|u\|_{H^2}^{1/2} \|g\|_{L^2}^{1/2} \|g\|_{H^1}^{1/2}.$$

Lemma 2.3 (Gronwall Inequality) *Let f be an absolutely continuous function that satisfies If $\frac{df}{dt} \leq gf + h$ for some real integrable functions $g(t)$ and $h(t)$. Then,*

$$f(t) \leq f(0) \exp \left(\int_0^t g(s) \, ds \right) + \int_0^t h(s) \exp \left(\int_s^t g(y) \, dy \right) ds.$$

3 Linearized and Adjoint Primitive Equations, and Differentiability

In this section, we prove regularity results for the linearized and adjoint primitive equations and establish the differentiability of the model solution with respect to the initial condition.

3.1 Linearized Equations and Differentiability

We linearize the primitive model equations around a strong solution $\psi = (\mathbf{v}, \theta)$ of (1). The resulting equations are also referred to as "tangent linear model." The linearized equations in terms of velocity and temperature variables $(\mathbf{V}, \Theta)$ are given by

$$\frac{\partial \mathbf{V}}{\partial t} + f\vec{k} \times \mathbf{V} + (\mathbf{V} \cdot \nabla)\mathbf{v} - \nabla \cdot \int_{-h}^z \mathbf{v}(x, y, \xi, t) \, d\xi \, \partial_z \mathbf{V}$$
$$+ (\mathbf{v} \cdot \nabla)\mathbf{V} - \nabla \cdot \int_{-h}^z \mathbf{V}(x, y, \xi, t) \, d\xi \, \partial_z \mathbf{v} - \nabla \left(\int_{-h}^z \Pi(x, y, \xi, t) \, d\xi + P_S \right)$$
$$- \frac{1}{Re_1} \triangle \mathbf{V} - \frac{1}{Re_2} \partial_{zz}^2 \mathbf{V} = F_{\mathbf{V}}, \tag{32a}$$

$$\nabla \cdot \mathbf{V} + \partial_z W = 0, \tag{32b}$$

$$\frac{\partial \Theta}{\partial t} + \nabla \cdot (\mathbf{V}\theta) + \partial_z(\mathbf{V}\theta) + \nabla \cdot (\mathbf{v}\Theta) + \partial_z(\mathbf{v}\Theta) - \frac{1}{Rt_1} \triangle\Theta - \frac{1}{Rt_2} \partial_{zz}^2 \Theta = F_\Theta, \tag{32c}$$

with surface pressure P_S and with the equation of state given by
$$\Pi := \rho_0 - \alpha(\Theta - \theta_0). \tag{32d}$$

The boundary conditions for (32) are analogous to the boundary condition (2)–(8) for the primitive equations (1) and read as follows:

$$on \; \Gamma_b := \{(x, y, z) \in \bar{\Omega} : z = -h\} : \; \mathbf{V} = 0, \; W = 0, \tag{33}$$
$$[\nabla_3 \Theta] \cdot n_3 = 0, \tag{34}$$

$$on \ \Gamma_s := \{(x, y, z) \in \bar{\Omega} : (x, y) \in \partial M\} : \mathbf{V} = 0, \ W = 0, \tag{35}$$

$$[\nabla_3 \Theta] \cdot \vec{n} = 0, \tag{36}$$

$$on \ \Gamma_u := \{(x, y, z) \in \bar{\Omega} : z = 0\} : \ \partial_z \mathbf{V} = 0, \ W = 0, \tag{37}$$

$$[\nabla_3 \Theta] \cdot n_3 = 0. \tag{38}$$

We write the linearized primitive equations for $\Psi = (V, \Theta)$ in the following form

$$\partial_t \Psi + \mathcal{B}'[\psi](\Psi) + L\Psi + D\Psi = F, \tag{39}$$

where D and L are defined in (24), (27), and where $\mathcal{B}'[\psi](\Psi) := (\mathcal{B}'_v[v](V), \mathcal{B}'_\theta[\theta](\Theta)$, with

$$\begin{aligned}
\mathcal{B}'_v[v](\mathbf{V}) &:= (\mathbf{v} \cdot \nabla \mathbf{V}) + (\mathbf{V} \cdot \nabla \mathbf{v}) - \left(\nabla \cdot \int_{-h}^z \mathbf{v}(x, y, \xi, t) \, d\xi \right) \partial_z \mathbf{V} \\
&\quad - \left(\nabla \cdot \int_{-h}^z \mathbf{V}(x, y, \xi, t) \, d\xi \right) \partial_z \mathbf{v}, \\
\mathcal{B}'_\theta[v](\Theta) &:= (\mathbf{v} \cdot \nabla \Theta) + (\mathbf{V} \cdot \nabla \theta) - \left(\nabla \cdot \int_{-h}^z \mathbf{v}(x, y, \xi, t) \, d\xi \right) \partial_z \Theta \\
&\quad - \left(\nabla \cdot \int_{-h}^z \mathbf{V}(x, y, \xi, t) \, d\xi \right) \partial_z \theta.
\end{aligned} \tag{40}$$

Theorem 3.1 (Regularity of linearized equations) *Let $\psi := (\mathbf{v}, \theta)$ be a strong solution of* (1) *on the time interval* $[0, T]$, $T > 0$. *Suppose $F := (F_\mathbf{V}, F_\Theta) \in L^2([0, T], \mathcal{L}^2(\Omega))$ is given. If the initial conditions for the linearized primitive equations satisfy $\Psi_0 := (\mathbf{V}_0, \Theta_0) \in \mathcal{V}$, then there exists a solution $\Psi := (\mathbf{V}, \Theta)$ of the linearized primitive equations* (32) *with the following properties,*

$$\begin{aligned}
\mathbf{V} &\in C([0, T], V_\mathbf{v}) \cap L^2([0, T], \mathbf{H}^2(\Omega)), \\
\Theta &\in C([0, T], V_\theta) \cap L^2([0, T], H^2(\Omega)), \\
\partial_t \mathbf{V} &\in L^2([0, T], \mathbf{L}^2(\Omega)), \\
\partial_t \Theta &\in L^2([0, T], L^2(\Omega)).
\end{aligned} \tag{41}$$

The solution Ψ satisfies for $t \in [0, T]$

$$\begin{aligned}
||\Psi(t)||_{\mathcal{H}^1}^2 &= ||\mathbf{V}(t)||_{\mathbf{H}^1}^2 + ||\Theta(t)||_{H^1}^2 \\
&\leq [||\mathbf{V}_0||_{\mathbf{H}^1}^2 + ||\Theta_0||_{H^1}^2] \exp \left\{ \int_0^t K_1(s) + K_2(s) \, ds \right\} \\
&\quad + c \int_0^t (||F_\mathbf{V}(s)||_{\mathbf{L}^2}^2 + ||F_\Theta(s)||_{L^2}^2) \exp \left\{ \int_0^s K_1(y) + K_2(y) \, dy \right\} ds,
\end{aligned} \tag{42}$$

and

$$\int_0^t ||\mathbf{V}(s)||_{\mathbf{H}^2}^2 + ||\Theta(s)||_{H^2}^2 \mathrm{d}s$$

$$\leq [||\mathbf{V}_0||_{\mathbf{H}^1}^2 + ||\Theta_0||_{H^1}^2] + \int_0^t (K_1(s) + K_2(s))(||\mathbf{V}(s)||_{\mathbf{H}^1}^2 + ||\Theta(s)||_{H^1}^2) \, \mathrm{d}s\}$$

$$+ \int_0^t (||F_{\mathbf{V}}(s)||_{\mathbf{L}^2}^2 + ||F_\Theta(s)||_{L^2}^2) \, \mathrm{d}s, \tag{43}$$

where

$$K_1(s) := ||\nabla \mathbf{v}||_{\mathbf{L}^2} ||\nabla \mathbf{v}||_{\mathbf{H}^1} + ||\mathbf{v}||_{\mathbf{H}^1}^4 + ||\nabla \, \partial_z \mathbf{v}||_{\mathbf{L}^2}^2 ||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 + 1$$
$$+ ||\nabla \, \mathbf{v}||_{\mathbf{L}^2}^2 ||\Delta \, \mathbf{v}||_{\mathbf{L}^2}^2 + ||\nabla \mathbf{v}||_{\mathbf{H}^1}^2 + ||\partial_z \nabla \mathbf{v}||_{\mathbf{L}^2}^4, \tag{44}$$
$$K_2(s) := ||\theta||_{H^1}^4 + ||\nabla \, \partial_z \theta||_{L^2}^2 ||\partial_z \theta||_{L^2}^2.$$

Proof The proof is given in Appendix 1.   □

The regularity result of the linearized primitive equations in Theorem 3.1 is instrumental for the differentiability result of the next Lemma.

Lemma 3.2 *Let $\psi_0 = (\mathbf{v}_0, \theta_0) \in \mathcal{V}$ and a time interval $[0, T]$, $T > 0$ be given. The mapping $\psi_0 \mapsto \psi(t; \psi_0)$ from $\mathcal{V}$ into $L^2([0, T], \mathcal{H}^1(\Omega))$ that assigns to an initial condition ψ_0 the solution $\psi(t; \psi_0) := \mathcal{M}[\psi_0](t)$ of the primitive equations (1) at time $t \in (0, T]$ has a Gateaux derivative $\frac{D\psi}{D\psi_0}\mathbf{a}$ in every direction $\mathbf{a} = (a_v, a_\theta) \in \mathcal{V}$. Furthermore, $\frac{D\psi}{D\psi_0}\mathbf{a}$ solves the linearized equations (32) with initial condition $\psi(t_0) = \mathbf{a}$ and forcing $F = (F_{\mathbf{V}}, F_\Theta) = 0$.*

Proof Let $\mathbf{a} := (a_v, a_\theta) \in \mathcal{V}$. Denote by $\psi_0, \psi_0 + \tau \mathbf{a} \in \mathcal{H}^1(\Omega)$, $\tau > 0$, two initial conditions and by ψ and $\psi_{\tau \mathbf{a}}$ the corresponding strong solutions of the primitive equations (1). They satisfy the equations

$$\frac{\partial \psi}{\partial t} + \mathcal{B}(\psi, \psi) + L\psi + D\psi = 0,$$

and

$$\frac{\partial \psi_{\tau \mathbf{a}}}{\partial t} + \mathcal{B}(\psi_{\tau \mathbf{a}}, \psi_{\tau \mathbf{a}}) + L\psi_{\tau \mathbf{a}} + D\psi_{\tau \mathbf{a}} = 0.$$

Let Ψ be the solution of the linearized equations (32), which are linearized around ψ,

$$\frac{\partial \Psi}{\partial t} + \mathcal{B}'[\psi](\Psi) + L\Psi + D\Psi = 0,$$

with zero forcing, initial condition $\Psi(t_0) = \mathbf{a}$ and boundary conditions (33)–(38). In order to prove the lemma, we have show that

$$y(\tau) := \psi_{\tau\mathbf{a}} - \psi - \tau\Psi$$
$$\text{satisfies} \ \lim_{\tau\to 0} \frac{||y(\tau)||_{L^2([0,T],\mathcal{H}^1)}}{|\tau|} = 0. \tag{45}$$

From the definition of y follows that it solves the equation

$$\frac{dy}{dt} + \mathcal{B}(\psi_{\tau h}, \psi_{\tau\mathbf{a}}) - \mathcal{B}(\psi, \psi) - \mathcal{B}'[\psi](\tau\Psi) + Ly + Dy = 0, \tag{46}$$

with initial condition $y_0 = 0$. We introduce $k = (k_u, k_\theta)$ as follows:

$$k := \mathcal{B}(\psi, \psi) - \mathcal{B}(\psi_{\tau\mathbf{a}}, \psi_{\tau\mathbf{a}}) + \mathcal{B}'[\psi](\psi_{\tau\mathbf{a}} - \psi).$$

The components of $k = (k_v, k_\theta)$ read as follows:

$$k_v := (\mathbf{v} \cdot \nabla)\mathbf{v} - (\mathbf{v}_{\tau\mathbf{a}} \cdot \nabla)\mathbf{v}_{\tau\mathbf{a}} + (\mathbf{v} \cdot \nabla)(\mathbf{v}_{\tau\mathbf{a}} - \mathbf{v}) + ((\mathbf{v}_{\tau\mathbf{a}} - \mathbf{v}) \cdot \nabla)\mathbf{v}$$
$$= ((\mathbf{v}_{\tau\mathbf{a}} - \mathbf{v}) \cdot \nabla)(\mathbf{v}_{\tau\mathbf{a}} - \mathbf{v}),$$
$$k_\theta := (\mathbf{v}_{\tau\mathbf{a}} - \mathbf{v}) \cdot \nabla(\theta_{\tau\mathbf{a}} - \theta).$$

Then, Eq. (46) for y becomes

$$\frac{dy}{dt} + \mathcal{B}'[\psi](y) + Ly + Dy = k. \tag{47}$$

Note that (47) is a linearized primitive equation with initial condition $y_0 = 0$ and an external forcing k that depends on the model solution itself. We prove now the following three inequalities: there exist constants $K, C > 0$ such that

$$
\begin{aligned}
&(i) \ \int_0^T ||y(t)||_{\mathcal{H}^1}^2 dt \leq K \int_0^T ||k||_{\mathcal{L}^2}^2 dt, \\
&(ii) \ ||k_v||_{\mathbf{L}^2} \leq C||\mathbf{v} - \mathbf{v}_{\tau a_v}||_{\mathbf{H}^2}||\mathbf{v} - \mathbf{v}_{\tau a_v}||_{\mathbf{H}^1} \\
&(iii) \ ||k_\theta||_{L^2} \leq C||\mathbf{v} - \mathbf{v}_{\tau a_\theta}||_{\mathbf{H}^2}||\theta_{\tau a_\theta} - \theta||_{H^1}.
\end{aligned}
\tag{48}
$$

For (i) follows from (175) in the proof of Theorem 3.1 (see Appendix 1) after integration with respect to time over a time interval $[0, t]$, with $t \in [0, T]$, with forcing $F_V = k_v$, $F_\Theta = k_\theta$, that

$$\int_0^t ||y(\tau)||_{\mathcal{H}^1}^2 \, d\tau \leq K(t) \int_0^t ||k_v(\tau)||_{\mathbf{L}^2}^2 + ||k_\theta||_{L^2}^2 d\tau.$$

where $K(t)$ is a bounded function on $[0, T]$. This proves assertion (i). For assertion (ii), we derive with the Agmon inequality

$$
\begin{aligned}
||k_v||_{\mathbf{L}^2} &= || \left((\mathbf{v}_{\tau a_v} - \mathbf{v}) \cdot \nabla \right) (\mathbf{v}_{\tau a_v} - \mathbf{v}) ||_{\mathbf{L}^2} \\
&\leq C ||\mathbf{v} - \mathbf{v}_{\tau a_v}||_{\mathbf{H}^2} ||\mathbf{v} - \mathbf{v}_{\tau a_v}||_{\mathbf{H}^1}.
\end{aligned}
\tag{49}
$$

Analogously, we obtain

$$
||k_\theta||_{L^2} \leq C ||\mathbf{v} - \mathbf{v}_{\tau a_v}||_{\mathbf{H}^2} ||\theta_{\tau a_\theta} - \theta||_{H^1}.
\tag{50}
$$

This proves (ii) and (iii). By combining (i)–(iii), we conclude that

$$
\int_0^T ||y(t)||_{\mathcal{H}^1}^2 \, dt \leq C \int_0^T \left(||\mathbf{v} - \mathbf{v}_{\tau a_v}||_{\mathbf{H}^1}^2 + ||\theta_{\tau a_\theta} - \theta||_{H^1}^2 \right) ||\mathbf{v}_{\tau a_v} - \mathbf{v}||_{\mathbf{H}^2}^2 \, dt.
\tag{51}
$$

Denote the difference between the two strong solutions by $\hat{\mathbf{v}} := \mathbf{v}_{\tau a_v} - \mathbf{v}$ and $\hat{\theta} := \theta_{\tau a_\theta} - \theta$. According to Theorem 1.1 we have $(\hat{\mathbf{v}}, \hat{\theta}) \in C([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega))$. Furthermore $(\hat{\mathbf{v}}, \hat{\theta})$ solves the equations

$$
\begin{aligned}
&\frac{\partial \hat{\mathbf{v}}}{\partial t} + (\hat{\mathbf{v}} \cdot \nabla)\mathbf{v}_{\tau a_v} + (\mathbf{v} \cdot \nabla)\hat{\mathbf{v}} + \hat{w}\partial_z \mathbf{v}_{\tau a_v} + w\partial_z \hat{\mathbf{v}} + \nabla \hat{p} + f\vec{k} \times \hat{\mathbf{v}} \\
&\quad - \frac{1}{Re_1}\triangle \hat{\mathbf{v}} - \frac{1}{Re_2}\partial_{zz}^2 \hat{\mathbf{v}} = 0, \\
&\partial_z \hat{p} - g\hat{\Pi} = 0, \\
&\nabla \cdot \mathbf{v} + \partial_z w = 0, \\
&\frac{\partial \hat{\theta}}{\partial t} + \hat{\mathbf{v}} \cdot \nabla \theta_{\tau a_\theta} + \mathbf{v} \cdot \nabla \hat{\theta} + \hat{w}\partial_z \theta_{\tau a_\theta} + w\partial_z \hat{\theta} - \frac{1}{Rt_1}\triangle \hat{\theta} - \frac{1}{Rt_2}\partial_{zz}^2 \hat{\theta}, \\
&\hat{\Pi} = -\alpha \hat{\Theta},
\end{aligned}
$$

with initial condition $\hat{\mathbf{v}}(t_0) = \tau a_v, \hat{\theta}(t_0) = \tau a_\theta$. This equation has a similar structure as the linearized equation (32) with a zero forcing term and homogeneous boundary conditions. Using the regularity properties of the two strong solutions $(\mathbf{v}_{\tau a_v}, \theta_{\tau a_\theta})$, $(\mathbf{v}, \theta)$, we can repeat all steps that lead to (175) (cf. Appendix 1), and this inequality implies

$$
\begin{aligned}
||\hat{\mathbf{v}}(t)||_{\mathbf{H}^1}^2 + ||\hat{\theta}(t)||_{H^1}^2 &\leq C(t) \left(||\hat{\mathbf{v}}_0||_{\mathbf{H}^1}^2 + ||\hat{\theta}_0||_{H^1}^2 \right) \\
&= C(t)\tau^2 \left(||a_v||_{\mathbf{H}^1}^2 + ||a_\theta||_{H^1}^2 \right),
\end{aligned}
\tag{52}
$$

with $C(t)$ bounded on $[0, T]$. Inserting this in (51) yields

$$
\int_0^T ||y(t)||_{\mathcal{H}^1}^2 \, dt \leq C(t)\tau^2 \left(||a_v||_{\mathbf{H}^1}^2 + ||a_\theta||_{H^1}^2 \right) \int_0^T ||\mathbf{v}_{\tau a_v} - \mathbf{v}||_{\mathbf{H}^2}^2 \, dt.
\tag{53}
$$

This proves (45), because $\mathbf{v}_{\tau a_v} - \mathbf{v} \in L^2([0, T], \mathbf{H}^2(\Omega)^2)$. With the definition of the Gateaux derivative it follows that the solution w of the linearized equations satisfies $w(\mathbf{a}) = (\mathcal{D}\Psi/\mathcal{D}\psi_0)\mathbf{a}$. $\square$

3.2 Adjoint Primitive Equations and Duality Relation

In this section, we study properties of the adjoint primitive equations. Adjoint equations allow to formulate a first-order necessary conditions for minima of the data assimilation cost functional and this condition provides the basis for the actual computation of minima within an optimization algorithm.

The adjoint equations are derived by choosing taking the $L^2([0, T], L^2(\Omega))$-scalar product of the linearized primitive equations with a so-called "*adjoint state variable*" $\widetilde{\Psi} := (\widetilde{\mathbf{V}}, \widetilde{\Theta}) := (\widetilde{U}, \widetilde{V}, \widetilde{\Theta})$. Then, all differential operators are moved from the linear variables to the adjoint variables by means of integration-by-parts. More details are given in Appendix 1 for more details. For an extensive treatment of adjoint equations, we refer to Marchuk et al. (1996).

The resulting adjoint primitive equations are given by

$$
\begin{aligned}
&- \partial_t \widetilde{U} + u \partial_x \widetilde{U} + \widetilde{U} \partial_x u + v \partial_y \widetilde{U} + \widetilde{U} \partial_x v + w \partial_z \widetilde{U} + \widetilde{W} \partial_z u + f \widetilde{U} \\
&+ \widetilde{U} \partial_x \theta + \theta \partial_x \widetilde{\Theta} + D_v(\widetilde{U}) = \tilde{F}_{\widetilde{U}}, \\
&- \partial_t \widetilde{V} + \widetilde{V} \partial_y u + v \partial_y \widetilde{V} + \widetilde{V} \partial_y v + u \partial_x \widetilde{V} + w \partial_z \widetilde{V} + \widetilde{W} \partial_z v - f \widetilde{V} \\
&+ \widetilde{V} \partial_y \theta + \theta \partial_y \widetilde{\Theta} + D_v(\widetilde{V}) = \tilde{F}_{\widetilde{V}}, \\
&\nabla \cdot \widetilde{\mathbf{V}} + \partial_z \widetilde{W} = 0, \\
&- \partial_t \widetilde{\Theta} + \mathbf{v} \cdot \nabla_3 \widetilde{\Theta} + \int_{-h}^z \nabla \cdot \widetilde{\mathbf{V}}(x, y, z') \, \mathrm{d}z' + D_\theta(\widetilde{\Theta}) = \tilde{F}_{\widetilde{\Theta}},
\end{aligned}
\tag{54}
$$

where $D_v(f) := -\frac{1}{Re_1} \triangle f - \frac{1}{Re_2} \partial_{zz}^2 f$ for $f \in \{\widetilde{U}, \widetilde{V}\}$, and $D_\theta(\widetilde{\Theta}) := -\frac{1}{Rt_1} \triangle \widetilde{\Theta} - \frac{1}{Rt_2} \partial_{zz}^2 \widetilde{\Theta}$. Furthermore does $\tilde{F} := (\tilde{F}_{\widetilde{U}}, \tilde{F}_{\widetilde{V}}, \tilde{F}_{\widetilde{\Theta}})$ denote a forcing term.

For $\widetilde{\Psi} := (\widetilde{U}, \widetilde{V}, \widetilde{\Theta})$, we write the adjoint primitive equations in the form

$$
- \partial_t \widetilde{\Psi} + \mathcal{B}^{'*}[\psi](\widetilde{\Psi}) + L \widetilde{\Psi} + D \widetilde{\Psi} = \tilde{F}, \tag{55}
$$

where $\mathcal{B}^{'*}[\psi](\widetilde{\Psi}) = \left(\mathcal{B}^{'*}[u](\widetilde{U}), \mathcal{B}^{'*}[v](\widetilde{V}), \mathcal{B}^{'*}[\theta](\widetilde{\Theta}) \right)$ is given by

$$
\begin{aligned}
\mathcal{B}^{'*}[\mathbf{v}](\widetilde{U}) &:= u \partial_x \widetilde{U} + \widetilde{U} \partial_x u + v \partial_y \widetilde{U} + \widetilde{U} \partial_x v + w \partial_z \widetilde{U} + \widetilde{W} \partial_z u, \\
\mathcal{B}^{'*}[\mathbf{v}](\widetilde{V}) &:= \widetilde{V} \partial_y u + v \partial_y \widetilde{V} + \widetilde{V} \partial_y v + u \partial_x \widetilde{V} + w \partial_z \widetilde{V} + \widetilde{W} \partial_z v, \\
\mathcal{B}^{'*}[\theta](\widetilde{\Theta}) &:= \mathbf{v} \cdot \nabla_3 \widetilde{\Theta}.
\end{aligned}
\tag{56}
$$

The operator $L\widetilde{\Psi} := (L_{\widetilde{U}} \widetilde{U}, L_{\widetilde{V}} \widetilde{V}, L_{\widetilde{\Theta}} \widetilde{\Theta})$ is defined as

$$
L_{\widetilde{U}} \widetilde{U} := f \widetilde{U} + \widetilde{U} \partial_x \theta + \widetilde{\Theta} \partial_x \theta,
$$

$$L_{\widetilde{V}}\widetilde{V} := -f\widetilde{V} + \widetilde{V}\partial_y\theta + \widetilde{\Theta}\partial_y\theta,$$

$$L_{\widetilde{\Theta}}\widetilde{\Theta} := \int_{-h}^{z} \nabla \cdot \widetilde{U}(x, y, z') \, \mathrm{d}z'. \tag{57}$$

The initial condition[2] of the adjoint equation is specified at $t = T$ and given by $(\widetilde{U}(t = T), \widetilde{V}(t = T), \widetilde{\Theta}(t = T)) = (\widetilde{U}_T, \widetilde{V}_T, \widetilde{\Theta}_T)$, the boundary conditions for (54) coincide with the corresponding boundary conditions (33)–(38) of the linearized equations and read as follows:

$$\text{on } \Gamma_b := \{(x, y, z) \in \bar{\Omega} : z = -h\} : \widetilde{\mathbf{V}} = 0, \widetilde{W} = 0, \tag{58}$$

$$[\nabla_3\widetilde{\Theta}] \cdot n_3 = 0, \tag{59}$$

$$\text{on } \Gamma_s := \{(x, y, z) \in \bar{\Omega} : (x, y) \in \partial M\} : \widetilde{\mathbf{V}} = 0, \tag{60}$$

$$\widetilde{W} = 0, [\nabla_3\widetilde{\Theta}] \cdot \vec{n} = 0, \tag{61}$$

$$\text{on } \Gamma_u := \{(x, y, z) \in \bar{\Omega} : z = 0\} : \partial_z\widetilde{\mathbf{V}} = 0, \widetilde{W} = 0, \tag{62}$$

$$[\nabla_3\widetilde{\Theta}] \cdot n_3 = 0. \tag{63}$$

The following result about the adjoint primitive equations follows immediately from the previous result on the linearized equations and the definition of the adjoint equations.

Theorem 3.3 (Regularity of Adjoint Equations) *Let $\psi := (v, \theta)$ be a strong solution of (1) on the time interval $[0, T]$, $T > 0$. Suppose $F := (F_{\mathbf{V}}, F_{\Theta}) \in L^2([0, T], \mathcal{L}^2(\Omega))$ is given. If the initial conditions for the adjoint linearized primitive equations, specified at $t = T$, satisfy $\widetilde{\Psi}_T := (\widetilde{\mathbf{V}}_T, \widetilde{\Theta}_T) \in \mathcal{V}$, then there exists a solution $\widetilde{\Psi} := (\widetilde{\mathbf{V}}, \widetilde{\Theta})$ of the adjoint linearized primitive equations (54) with the following properties:*

$$\begin{aligned} &\widetilde{\mathbf{V}} \in C([0, T], V_{\mathbf{v}}) \cap L^2([0, T], \mathbf{H}^2(\Omega)), \\ &\widetilde{\Theta} \in C([0, T], V_\theta) \cap L^2([0, T], H^2(\Omega)), \\ &\partial_t\widetilde{\mathbf{V}} \in L^2([0, T], \mathbf{L}^2(\Omega)), \\ &\partial_t\widetilde{\Theta} \in L^2([0, T], L^2(\Omega)). \end{aligned} \tag{64}$$

The solution $\widetilde{\Psi}$ satisfies for $t \in [0, T]$

$$\begin{aligned} \|\widetilde{\Psi}(t)\|_{\mathcal{H}^1} &= \|\widetilde{\mathbf{V}}(t)\|_{\mathbf{H}^1}^2 + \|\widetilde{\Theta}(t)\|_{H^1}^2 \\ &\leq [\|\widetilde{\mathbf{V}}_T\|_{\mathbf{H}^1}^2 + \|\widetilde{\Theta}_T\|_{H^1}^2] \exp\{\int_t^T K_1(s) + K_2(s) \, \mathrm{d}s\} \\ &+ c \int_t^T (\|\widetilde{F}_{\widetilde{\mathbf{V}}}(s)\|_{\mathbf{L}^2}^2 + \|\widetilde{F}_{\widetilde{\Theta}}(s)\|_{L^2}^2) \exp\{\int_0^s K_1(y) + K_2(y) \, \mathrm{d}y\} \, \mathrm{d}s, \end{aligned} \tag{65}$$

---

[2] Adjoint equations evolve backward in time, starting at time $t = T$. In integrals of adjoint quantities over time we use the convention that the lower integration limit is the smaller number and the upper integration limit is the larger number, i.e., we integrate always from earlier to later.

and

$$\int_t^T ||\widetilde{\mathbf{V}}(s)||_{\mathbf{H}^2}^2 + ||\widetilde{\Theta}(s)||_{H^2}^2 \mathrm{d}s$$

$$\leq [||\widetilde{\mathbf{V}}_0||_{\mathbf{H}^1}^2 + ||\widetilde{\Theta}_0||_{H^1}^2] + \int_0^t (K_1(s) + K_2(s))(||\widetilde{\mathbf{V}}(s)||_{\mathbf{H}^1}^2 + ||\widetilde{\Theta}(s)||_{H^1}^2)\,\mathrm{d}s\}$$

$$+ \int_t^T (||\tilde{F}_{\widetilde{\mathbf{V}}}(s)||_{\mathbf{L}^2}^2 + ||\tilde{F}_{\widetilde{\Theta}}(s)||_{L^2}^2)\,\mathrm{d}s, \tag{66}$$

where

$$K_1(s) := ||\nabla \mathbf{v}||_{\mathbf{L}^2}||\nabla \mathbf{v}||_{\mathbf{H}^1} + ||\mathbf{v}||_{\mathbf{H}^1}^4 + ||\nabla \partial_z \mathbf{v}||_{\mathbf{L}^2}^2||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 + 1$$
$$+ ||\nabla \mathbf{v}||_{\mathbf{L}^2}^2||\Delta \mathbf{v}||_{\mathbf{L}^2}^2 + ||\nabla \mathbf{v}||_{\mathbf{H}^1}^2 + ||\partial_z \nabla \mathbf{v}||_{\mathbf{L}^2}^4,$$
$$K_2(s) := ||\theta||_{H^1}^4 + ||\nabla \partial_z \theta||_{L^2}^2||\partial_z \theta||_{L^2}^2. \tag{67}$$

Proof The proof follows essentially from the corresponding result on the linearized equations, i.e., Theorem 3.1, and the duality relation between linearized and adjoint equations. □

The next lemma follows from the definition of linearized and adjoint equations and integration by parts.

Lemma 3.4 (Adjoint Relation) *Let $F, \tilde{F} \in L^2([0, T], \mathcal{L}^2(\Omega))$ be the forcing of the linearized equations and the adjoint equations, respectively. By Ψ and $\tilde{\Psi}$, we denote the variables of the linear and adjoint equations. Then,*

$$\langle F, \tilde{\Psi} \rangle_{L^2([0,T]\mathcal{L}^2)} = \langle \tilde{F}, \Psi \rangle_{L^2([0,T]\mathcal{L}^2)} - \int_\Omega \Psi \cdot \tilde{\Psi}|_0^T\,\mathrm{d}x.$$

4 Solvability and Computation of the Data Assimilation Problem

4.1 Existence of Local Minima of the Data Assimilation Problem

Based on the regularity results of the previous section, we formulate now the specific data assimilation cost functional for the primitive equations. The model dynamics that we are considering consist of trajectories in $C([0, T], \mathcal{H}^1(\Omega))$ and are emerging from initial conditions in $\mathcal{H}^1(\Omega)$. Let observations $\psi_{\mathrm{obs}} \in C([0, T], \mathcal{H}^1(\Omega))$ be given that reside in the same space as the dynamics of the ocean primitive equations. Suppose furthermore that a background guess $\psi_{\mathrm{back}} \in \mathcal{H}^1(\Omega)$ at initial time is given. We define the cost functional by

$$\mathcal{J}(\psi_0) := \mathcal{J}_b(\psi_0) + \mathcal{J}_{\mathrm{obs}}(\psi_0). \tag{68}$$

The background term is given by

$$
\begin{aligned}
\mathcal{J}_b(\psi_0) &:= \big\langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \psi_0 - \psi_{\text{back}} \big\rangle_{\mathcal{H}^1} \\
&= \sum_{|\alpha| \le 1} \int_\Omega \mathcal{D}^\alpha \mathcal{B}(\psi_0 - \psi_{\text{back}}) \cdot \mathcal{D}^\alpha (\psi_0 - \psi_{\text{back}}) \, \mathrm{d}x \mathrm{d}y \mathrm{d}t,
\end{aligned}
\tag{69}
$$

where $\mathcal{D}^\alpha$ denotes a derivative with multi-index α. The observational term is defined as

$$
\begin{aligned}
\mathcal{J}_{\text{obs}}(\psi_0) &:= \int_0^T \big\langle \mathcal{R}(\mathcal{M}[\psi_0](t) - \psi_{\text{obs}}(t)), \mathcal{M}[\psi_0](t) - \psi_{\text{obs}}(t) \big\rangle_{\mathcal{H}^1} \mathrm{d}t \\
&= \sum_{|\alpha| \le 1} \int_0^T \int_\Omega \mathcal{D}^\alpha \mathcal{R} \left(\mathcal{M}[\psi_0](t) - \psi_{\text{obs}}(t)\right) \cdot \mathcal{D}^\alpha \left(\mathcal{M}[\psi_0](t) - \psi_{\text{obs}}(t)\right) \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t.
\end{aligned}
\tag{70}
$$

The model error covariance operator $\mathcal{B}$ provides a weighting according to the estimated background error and is assumed to be linear, bounded and positive definite. We do not address here the important problem of modeling error statistics but assume these error covariance operators as given. Furthermore, we assume that the error covariance operators preserve the functional space of model solutions and observations, i.e., we make the

General Assumption on $\mathcal{B}, \mathcal{R}: F \in \mathcal{X}$ implies $\mathcal{B}F \in \mathcal{X}$ and $\mathcal{R}F \in \mathcal{X}$.   (71)

We note that the notion of strong solution, given in Definition 2.1, contains the $L^2([0, T], \mathcal{H}^2(\Omega)$-norm for the regularity of solutions to the primitive equations, while the observational part $\mathcal{J}_{\text{obs}}$ of our cost functional in (70) imposes only the weaker $L^2([0, T], \mathcal{H}^1(\Omega)$-norm. The reason for this is the adjoint characterization of optimal initial conditions in Theorem 4.4. Details are explained in Remark 2 after Theorem 4.4 is stated. The proof of Theorem 4.1 does also work if the stronger $L^2([0, T], \mathcal{H}^2(\Omega)$-norm are used in $\mathcal{J}_{\text{obs}}$. In the proof, we use the regularity that is given by the well-posedness Theorem 1.1 and not the regularity that is implied by the boundedness of the observational part of the cost functional. We state this as a separate results in Corollary 4.2.

Theorem 4.1 (Optimal Initial Conditions) *Let observations $\psi_{\text{obs}} \in C([0, T], \mathcal{V})$ and a background guess $\psi_{\text{back}} \in \mathcal{H}^1(\Omega)$ be given. Then, there exist minimizers $\bar\psi_0 = (\mathbf{v}_0, \theta_0) \in \mathcal{V}$ of the data assimilation problem (9) with cost functional given by (68). We refer to these minimizers also as optimal initial conditions for the primitive equations (1).*

Proof Let $(\psi_{0,n})_n = (\mathbf{v}_{0,n}, \theta_{0,n})_n \subseteq \mathcal{V} \subseteq \mathcal{H}^1(\Omega)$ be a minimizing sequence of initial conditions for the data assimilation problem. We denote by $\psi_n := \mathcal{M}[\psi_{0,n}] \in C([0, T], \mathcal{V}) \cap L^2([0, T], \mathcal{H}^2(\Omega))$ the corresponding solutions of the primitive equations (1). The model-observation difference satisfies $(\mathcal{M}[\psi_0] - \psi_{\text{obs}}) \in C([0, T], \mathcal{H}^1(\Omega))$. Since the model error covariance operator $\mathcal{R}$ preserves the space

(cf. (71)), it follows that $\mathcal{R}(\mathcal{M}[\psi_0] - \psi_{\mathrm{obs}}) \in C([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega))$, and from (70), we infer that the cost functional is well-defined. (68) follows that the sequence of initial conditions $(\psi_{0,n})_n$ is bounded in $\mathcal{H}^1(\Omega)$, i.e., there exists a $c > 0$ such that uniformly for all $n \in \mathbb{N}$

$$||\psi_{0,n}||_{\mathcal{H}^1} \leq c. \tag{72}$$

From Theorem 1.1, we see that the sequence of associated solutions $(\psi_n)_n = (v_n, \theta_n)_n$ is bounded in $C([0, T], \mathcal{V}) \cap L^2([0, T], \mathcal{H}^2(\Omega))$, in particular there exists a $C > 0$ such that for all $n \in \mathbb{N}$

$$\sup_{t \in [0,T]} ||\psi_n(t)||_{H^1} \leq C \quad \text{and} \quad \int_0^T ||\psi_n(t)||_{\mathcal{H}^2}^2 \mathrm{d}t \leq C. \tag{73}$$

Since $H^2(\Omega)$ is compactly embedded in $H^1(\Omega)$, we conclude that a subsequence, still denoted $(\psi_n)_n = (v_{0,n}, \theta_{0,n})_n$, exists and a limit $\bar{\psi} = (\bar{v}, \bar{\Theta})$, such that $(\psi_n)_n$ converges to $\bar{\psi}$ weakly in $L^2([0, T], \mathcal{H}^2(\Omega))$ and strongly in $L^2([0, T], \mathcal{H}^1(\Omega))$.

We show now that the limit $\bar{\psi}(t)$ is a strong solution of the primitive equations, i.e., $\bar{\psi} = \mathcal{M}[\bar{\psi}_0]$. Consider the equation (19) in Definition 2.1 of strong solutions of (1) for the elements of the sequence $\psi_n = (v_n, \theta_n)_n$ with initial conditions $(\psi_{0,n})_n = (v_{0,n}, \theta_{0,n})_n$. We integrate over the time interval $[0, T]$, use that $(\partial_t v_n, \partial_t \theta_n) \in L^2([0, T], \mathcal{L}^2(\Omega))$, and obtain for all $\Phi \in \mathbf{C}^\infty(\Omega)^2$ and $\phi \in C^\infty(\Omega)$

$$\int_\Omega (v_n(T) - v_n(0)) \cdot \Phi \, \mathrm{d}x + \int_0^T \int_\Omega [(v_n \cdot \nabla) v_n] \cdot \Phi \, \mathrm{d}x \mathrm{d}t$$

$$+ \int_0^T \int_\Omega f \vec{k} \times v_n \cdot \Phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$- \int_0^T \int_\Omega \left(\int_{-h}^z v_n(x, y, \xi, t) \, \mathrm{d}\xi \right) \cdot \partial_z v_n \, (\nabla \cdot \Phi) \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$+ \int_0^T \int_{\Gamma_u} (\nabla p_{s,n}) \cdot \Phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$- \int_0^T \int_\Omega \left(\nabla \int_{-h}^z \rho(\theta_n)(x, y, \xi, t) \, \mathrm{d}\xi \right) \cdot \Phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$+ \frac{1}{Re_1} \int_0^T \int_\Omega \nabla v_n \cdot \nabla \Phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t + \frac{1}{Re_2} \int_0^T \int_\Omega \partial_v v_n \cdot \partial_z \Phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t = 0,$$

$$\nabla \cdot v_n + \partial_z w_n = 0,$$

$$\int_0^T \int_\Omega \partial_t \theta_n \phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t + \int_0^T \int_\Omega (v_n \cdot \nabla) \theta_n \phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$- \int_0^T \int_\Omega \left(\nabla \cdot \int_{-h}^0 v_n(x, y, \xi, t) \, \mathrm{d}\xi \right) (\partial_z \theta_n) \phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$+ \frac{1}{Rt_1} \int_0^T \int_\Omega \nabla_3 \theta_n \cdot \nabla_3 \phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t + \frac{1}{Rt_2} \int_0^T \int_\Omega \partial_z \theta_n \partial_z \phi \, \mathrm{d}x \mathrm{d}y \mathrm{d}z \mathrm{d}t$$

$$+ k_\theta \int_0^T \int_{\Gamma_u} \theta_n \phi \, dx dy dt = 0, \tag{74a}$$

where $\rho(\theta_n) = \rho_0 - \alpha(\theta_n - \theta_0)$ (cf. (1e)). Since $(\psi_n)_n$ converges strongly in $L^2([0, T], \mathcal{H}^1(\Omega))$ to $\bar\psi \in L^2([0, T], \mathcal{H}^1(\Omega))$ there exists a subset $[0, T'] \subset [0, T]$, such that $[0, T] \setminus [0, T']$ has Lebesgue measure zero and $(v_n(t), \theta_n(t))_n$ converges to $(\bar v(t), \bar\theta(t))$ in $\mathcal{H}^1$ for all $t \in [0, T']$. This implies for almost every $t_0, t_1 \in [0, T]$ as $n \to \infty$

$$\int_\Omega (\mathbf{v}_n(t_1) - \mathbf{v}_n(t_0) - \bar{\mathbf{v}}(t_1) + \bar{\mathbf{v}}(t_0)) \cdot \Phi \, dx dy dz \to 0. \tag{75}$$

With the convergence of $(\psi_n)_n$ to $\bar\psi$ in $L^2([0, T], \mathcal{H}^1(\Omega))$ the following convergence properties follows

$$\int_0^T \int_\Omega f\vec{k} \times (\mathbf{v}_n - \bar{\mathbf{v}}) \cdot \Phi \, dx dy dz dt \to 0,$$

$$\int_0^T \int_\Omega (\nabla \int_{-h}^z \big(\rho(\theta_n)(x, y, \xi, t) - \rho(\bar\theta)(x, y, \xi, t)\big) \, d\xi) \cdot \Phi \, dx dy dz dt \to 0,$$

$$\frac{1}{Re_1} \int_0^T \int_\Omega \nabla(\mathbf{v}_n - \bar{\mathbf{v}}) \cdot \nabla \Phi \, dx dy dz dt \to 0,$$

$$\frac{1}{Re_2} \int_0^T \int_\Omega \partial_z(\mathbf{v}_n - \bar{\mathbf{v}}) \cdot \partial_z \Phi \, dx dy dz dt \to 0.$$

For the nonlinear term in the velocity equation (74a), we obtain after integration by parts and with the inequalities of Cauchy–Schwartz, Hölder and Ladyshenzkaya (31),

$$\left| \int_0^T \int_\Omega [(\mathbf{v}_n \cdot \nabla)\mathbf{v}_n + w_n \partial_z \mathbf{v}_n - (\bar{\mathbf{v}} \cdot \nabla)\bar{\mathbf{v}} - \bar w \partial_z \bar{\mathbf{v}}] \cdot \Phi \, dx dy dz dt \right|$$

$$= \left| \int_0^T \int_\Omega [((\mathbf{v}_n - \bar{\mathbf{v}}) \cdot \nabla)\mathbf{v}_n + (w_n - \bar w)\partial_z \mathbf{v}_n] \cdot \Phi \, dx dy dz dt \right.$$

$$\left. + \int_0^T \int_\Omega [\bar{\mathbf{v}} \cdot \nabla)\Phi] \cdot (\mathbf{v}_n - \bar{\mathbf{v}}) + [\bar w \partial_z \Phi] \cdot (\mathbf{v}_n - \bar{\mathbf{v}}) \, dx dy dz dt \right|$$

$$\leq c \int_0^T ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{L}^6} ||\nabla \mathbf{v}_n||_{L^2} ||\Phi||_{\mathbf{L}^3} dt + c \int_0^T ||w_n - \bar w||_{L^2} ||\partial_z \mathbf{v}_n||_{\mathbf{L}^6} ||\Phi||_{\mathbf{L}^3} dt$$

$$+ c \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{L}^3} ||\nabla \Phi||_{\mathbf{L}^2} ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{L}^6} dt + c \int_0^T ||\bar w||_{\mathbf{L}^3} ||\partial_z \Phi||_{\mathbf{L}^2} ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{L}^6} dt$$

$$\leq c \int_0^T ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{H}^1} ||\nabla \mathbf{v}_n||_{\mathbf{L}^2} ||\Phi||_{\mathbf{H}^1} dt + c \int_0^T ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{H}^1} ||\partial_z \mathbf{v}_n||_{\mathbf{H}^1} ||\Phi||_{\mathbf{H}^1} dt$$

$$+ c \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^1} ||\nabla \Phi||_{\mathbf{L}^2} ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{H}^1} dt + c \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^2} ||\partial_z \Phi||_{\mathbf{L}^2} ||\mathbf{v}_n - \bar{\mathbf{v}}||_{\mathbf{H}^1} dt$$

$$\leq c ||\mathbf{v}_n - \bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^1)} ||\mathbf{v}_n||_{L^2([0,T],\mathbf{H}^1)} ||\Phi||_{\mathbf{H}^1}$$

$$+ c||\mathbf{v}_n - \bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^1)} ||\mathbf{v}_n||_{L^2([0,T],\mathbf{H}^1)} ||\Phi||_{\mathbf{H}^1}$$
$$+ c||\bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^1)} ||\nabla\Phi||_{\mathbf{L}^2} ||\mathbf{v}_n - \bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^1)}$$
$$+ c||\bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^2)} ||\partial_z\Phi||_{\mathbf{L}^2} ||\mathbf{v}_n - \bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^1)}. \tag{76}$$

The first three terms on the right-hand side of (76) vanish due to the convergence of $(\mathbf{v}_n)_n$ to $\bar{\mathbf{v}}$ in $L^2([0,T],\mathcal{H}^1(\Omega))$ and the boundedness of the remaining terms in the respective integral. For the last term in (76), we note that due to (73), it holds that $||\bar{\mathbf{v}}||_{L^2([0,T],\mathbf{H}^2)} \leq \liminf_{n\to\infty} ||\mathbf{v}_n||_{L^2([0,T],\mathbf{H}^2)} < C$. Since $\bar{\mathbf{v}} \in L^2([0,T],\mathbf{H}^2)$ the convergence of $(\mathbf{v}_n)_n$ in $L^2([0,T],H^1(\Omega)$ shows that the last term converges to zero. The convergence of the nonlinear terms in the temperature equations follows analogously. The pointwise convergence of $(\mathbf{v}_n)_n$ to $\bar{\mathbf{v}}$ in $\mathbf{H}^1$ for almost every $t \in [0,T]$ implies that $\bar{\mathbf{v}}$ satisfies the divergence-free constraint for almost every $t \in [0,T]$. This proves that the limit $\bar{\psi}$ is a solution of the primitive equations.

It remains to show that that $\bar{\psi} \in C([0,T],\mathcal{V})$. We show first that $\partial_t\bar{\psi} \in L^2([0,T],\mathcal{L}^2(\Omega))$. In order to establish this assertion, we take the L^2-scalar product of the primitive equations (1a)–(1e) for $\psi = \bar{\psi}$ with $\partial_t\bar{\psi} = \partial_t(\bar{\mathbf{v}},\bar{\theta})$. For the velocity equation, we obtain

$$\int_0^T \int_\Omega |\partial_t\bar{\mathbf{v}}|^2 \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t$$
$$\leq \left| \int_0^T \int_\Omega [(\bar{\mathbf{v}}\cdot\nabla)\bar{\mathbf{v}}]\cdot\partial_t\bar{\mathbf{v}} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \right| + \left| \int_0^T \int_\Omega \left[\left(\int_0^z \nabla\cdot\bar{\mathbf{v}}\,\mathrm{d}\xi \right) \partial_z\bar{\mathbf{v}} \right]\cdot\partial_t\bar{\mathbf{v}} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \right|$$
$$+ \left| \int_0^T \int_\Omega \nabla\left(\int_{-h}^z g\bar{\rho}(x,y,\xi,t)\,\mathrm{d}\xi + \bar{p}_S \right)\cdot\partial_t\bar{\mathbf{v}} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \right|$$
$$+ \left| \int_0^T \int_\Omega \frac{1}{Re_1}\triangle\bar{\mathbf{v}}\cdot\partial_t\bar{\mathbf{v}} + \frac{1}{Re_2}\partial_{zz}^2\bar{\mathbf{v}}\cdot\partial_t\bar{\mathbf{v}} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \right|, \tag{77}$$

where $\bar{\rho}$ is the density computed from $\bar{\theta}$ and where we omitted the external forcing for simplicity. We consider the nonlinear terms on the right-hand side. With the inequalities of Hölder, Ladyzhenskaya and Young follows

$$\left| \int_0^T \int_\Omega [(\bar{\mathbf{v}}\cdot\nabla)\bar{\mathbf{v}}]\cdot\partial_t\bar{\mathbf{v}} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \right| \leq \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{L}^3} ||\nabla\bar{\mathbf{v}}||_{\mathbf{L}^6} ||\partial_t\bar{\mathbf{v}}||_{\mathbf{L}^2} \mathrm{d}t$$
$$\leq \frac{c}{2\epsilon_1} \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^1}^2 ||\nabla\bar{\mathbf{v}}_{\mathbf{H}^1}||^2 \mathrm{d}t + \frac{\epsilon_1}{2} \int_0^T ||\partial_t\bar{\mathbf{v}}||_{\mathbf{L}^2}^2, \tag{78}$$

where $\epsilon_1 > 0$ is arbitrary. For the vertical advection follows with Lemma 2.2 and the Young inequality

$$\left| \int_0^T \int_\Omega \left[\left(\int_0^z \nabla \cdot \bar{\mathbf{v}} \, d\xi \right) \partial_z \bar{\mathbf{v}} \right] \cdot \partial_t \bar{\mathbf{v}} \, dx dy dz dt \right|$$

$$\leq \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^1}^{1/2} ||\bar{\mathbf{v}}||_{\mathbf{H}^2}^{1/2} ||\partial_z \bar{\mathbf{v}}||_{\mathbf{H}^1}^{1/2} ||\partial_z \bar{\mathbf{v}}||_{\mathbf{L}^2}^{1/2} ||\partial_t \bar{\mathbf{v}}||_{\mathbf{L}^2} dt \tag{79}$$

$$\leq \frac{c}{2\epsilon_2} \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^1} ||\bar{\mathbf{v}}||_{\mathbf{H}^2} ||\partial_z \bar{\mathbf{v}}||_{\mathbf{H}^1} ||\partial_z \bar{\mathbf{v}}||_{\mathbf{L}^2} + \frac{\epsilon_2}{2} |\int_0^T |\partial_t \bar{\mathbf{v}}||_{\mathbf{L}^2}^2 dt,$$

with $\epsilon_2 > 0$. The remaining terms in (77) can be bounded with the inequality of Cauchy–Schwarz and Young. In each of these upper bounds, a term occurs that involves the L^2-norm of the time derivative $\partial_t \bar{\mathbf{v}}$ multiplied by an ϵ_i as in (78) and (79). If we choose the ϵ_i appropriately, we can compensate the time derivative on the right-hand side and the time derivative on the left-hand side. In summary, we obtain the inequality

$$\int_0^T |||\partial_t \bar{\mathbf{v}}||_{\mathbf{L}^2}^2 dt \leq c \int_0^T ||\bar{\theta}||_{H^1}^2 + ||\bar{p}_s||_{H^1}^2 dt + c \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^2}^2 dt$$

$$+ c \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^1}^2 ||\nabla \bar{\mathbf{v}}_{\mathbf{H}^1}||^2 dt + c \int_0^T ||\bar{\mathbf{v}}||_{\mathbf{H}^1} ||\bar{\mathbf{v}}||_{\mathbf{H}^2} ||\partial_z \bar{\mathbf{v}}||_{\mathbf{H}^1} ||\partial_z \bar{\mathbf{v}}||_{\mathbf{L}^2} dt. \tag{80}$$

Using the regularity $\bar{\psi} = (\bar{\mathbf{v}}, \bar{\theta}) \in L^2([0, T], \mathcal{H}^2(\Omega))$, we observe that the right-hand side of (80) is finite and this shows that $\partial_t \bar{\mathbf{v}} \in L^2([0, T], \mathbf{L}^2(\Omega))$. From the same arguments follows for the temperature equation that $\partial_t \bar{\theta} \in L^2([0, T], L^2(\Omega))$. Since $\bar{\psi} \in L^2([0, T], \mathcal{H}^2(\Omega))$, $\partial_t \bar{\psi} \in L^2([0, T], \mathcal{L}^2(\Omega))$ and $\mathcal{H}^2 \subseteq \mathcal{H}^1 \subseteq \mathcal{L}^2$ are compactly embedded it follows with the Lions–Aubins compactness lemma that $\bar{\psi} \in C([0, T], \mathcal{H}^1(\Omega))$. This implies that (75) as well as the divergence-free constraint are satisfied for all $t \in [0, T]$. This proves that the limit $\bar{\psi}$ is a solution of the primitive equations in the sense of Definition 2.1.

We show now that the initial condition $\bar{\psi}_0$ minimizes the cost functional $\mathcal{J}$ in (68). Lower semi-continuity of the scalar product implies for the limit $\bar{\psi}_0$

$$\langle \mathcal{B}(\bar{\psi}_0 - \psi_{\text{back}}), \bar{\psi}_0 - \psi_{\text{back}} \rangle_{\mathcal{H}^1} \leq \liminf_n \langle \mathcal{B}(\psi_{0,n} - \psi_{\text{back}}), \psi_{0,n} - \psi_{\text{back}} \rangle_{\mathcal{H}^1}, \tag{81}$$

and for the associated sequence of model solutions

$$\int_0^T \langle \mathcal{R}(\mathcal{M}[\bar{\psi}_0] - \psi_{\text{obs}}), \mathcal{M}[\bar{\psi}_0] - \psi_{\text{obs}} \rangle_{\mathcal{H}^1} dt$$

$$\leq \liminf_n \int_0^T \langle \mathcal{R}(\mathcal{M}[\psi_{0,n}] - \psi_{\text{obs}}), \mathcal{M}[\psi_{0,n}] - \psi_{\text{obs}} \rangle_{\mathcal{H}^1} dt. \tag{82}$$

As a consequence of (81) and (82), we have

$$\mathcal{J}(\bar{\psi}_0) \leq \liminf_n \mathcal{J}(\psi_{0,n}). \tag{83}$$

The initial condition $\bar{\psi}_0$ of $\bar{\psi}$ is a minimizer of $\mathcal{J}$, defined in (68).   □

From the proof of Theorem 4.1 follows immediately the next corollary that replaces the $\mathcal{H}^1$-scalar product in the observational part of the cost functional by the H^2-scalar product. This choice reflects the regularity of the solutions of the primitive equations. With regard to this aspect, see also Remark 2 below.

Corollary 4.2 *Let observations* $\psi_{\mathrm{obs}} \in L^2([0, T], \mathcal{H}^2(\Omega))$ *and a background guess* $\psi_{\mathrm{back}} \in \mathcal{H}^1(\Omega)$ *be given. Let the observational term* $\mathcal{J}_{\mathrm{obs}}$ *in* (70) *of the cost functional* (68) *be defined by the* $\mathcal{H}^2$-*scalar product*

$$\mathcal{J}_{\mathrm{obs}}(\psi_0) := \int_0^T \left\langle \mathcal{R}(\mathcal{M}[\psi_0](t) - \psi_{\mathrm{obs}}(t)), \mathcal{M}[\psi_0](t) - \psi_{\mathrm{obs}}(t) \right\rangle_{\mathcal{H}^2} \mathrm{d}t. \quad (84)$$

Then, there exist minimizers $\bar{\psi}_0 = (\mathbf{v}_0, \theta_0) \in \mathcal{H}^1(\Omega)$ *of the data assimilation problem* (9) *with cost functional given by* (68) *and background term defined in* (69) *and observational term in* (84).

4.2 Adjoint Characterization of Local Minima

The existence of local minima of the data assimilation problem is guaranteed by Theorem 4.1. We address now the problem of how these optimal initial conditions can be computed. For this purpose, we prove an necessary condition in terms of the adjoint equation. As preparation we need the following Lemma on the regularity of Helmholtz equations for temperature and velocity.

Lemma 4.3 *(i) Let* $f \in L^2(\Omega)$ *be given. The equation*

$$(Id - \triangle_3)\theta = f, \qquad (\triangle_3 := \partial_{xx}^2 + \partial_{yy}^2 + \partial_{zz}^2) \quad (85)$$

with homogeneous Neumann boundary condition $\nabla_3 \theta \cdot \vec{n} = 0$ *at* $\partial\Omega$ *has a unique solution* $\theta \in H^2(\Omega)$ *that satisfies*

$$||\theta||_{H^2} \leq c||f||_{L^2}. \quad (86)$$

(ii) Let $F \in \mathbf{L}^2(\Omega)^2$ *be given. The equation*

$$(Id - \Delta_3)\mathbf{v} = F, \qquad (\Delta_3 := \partial_{xx}^2 + \partial_{yy}^2 + \partial_{zz}^2), \quad (87)$$

with mixed homogeneous boundary conditions

$$\mathbf{v} = 0, \text{ on } \Gamma_s \cup \Gamma_b \quad and \quad \partial_z \mathbf{v} = 0 \text{ on } \Gamma_u, \quad (88)$$

has a unique solution $\mathbf{v} \in \mathbf{H}^2\Omega$ *that satisfies*

$$||\mathbf{v}||_{\mathbf{H}^2} \leq c||F||_{\mathbf{L}^2}. \quad (89)$$

Proof For equation (85) with homogeneous Neumann boundary conditions we refer to Theorem 9.26 in Brezis (2010) from which the assertion follows. For equation (87) and boundary condition (88), we note that the associated bilinearform

$$b(\mathbf{v}_1, \mathbf{v}_2) := \int_\Omega \mathbf{v}_1 \cdot \mathbf{v}_2 \mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega \nabla\mathbf{v}_1 \cdot \nabla\mathbf{v}_2 \mathrm{d}x\mathrm{d}y\mathrm{d}z, \quad \mathbf{v}_1, \mathbf{v}_2 \in V_\mathbf{v}$$

satisfies $b(\mathbf{v}, \mathbf{v}) \geq ||\mathbf{v}||^2_{\mathbf{H}^1}$ for $\mathbf{v} \in V_\mathbf{v}$. The assertion follows from a classical result on elliptic equations (see Necas 2012, Theorem 3.1. on p. 30). $\qquad\square$

Theorem 4.4 (*Adjoint Characterization of Optimal Initial Conditions*) *Let observations* $\psi_{\mathrm{obs}} \in C([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega))$ *and a background state* $\psi_{\mathrm{back}} \in \mathcal{H}^1(\Omega)$ *be given. Denote by* $\bar{\psi}_0 = (\bar{\mathbf{v}}_0, \bar{\Theta}_0) \in \mathcal{H}^1(\Omega)$ *an optimal initial condition of the data assimilation problem* (9). *Then* $\bar{\psi}_0$ *satisfies*

$$\bar{\psi}_0 = \psi_{\mathrm{back}} - \mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}_0, \tag{90}$$

where $\tilde{\Psi}$ *is the solution of the adjoint equation* (54) *with initial condition* $\tilde{\Psi}(T) = 0$, *specified at time* $t = T$ *and with forcing given by*

$$\tilde{F} := \mathcal{S}\mathcal{R}(\mathcal{M}[\bar{\psi}_0] - \psi_{\mathrm{obs}}). \tag{91}$$

The operator $\mathcal{S} := (\mathcal{S}_\mathbf{v}, \mathcal{S}_\theta)$, *with* $\mathcal{S}_\mathbf{v} := (Id - \Delta_3)$ *defined in* (87) *and boundary conditions* (88), *and* $\mathcal{S}_\theta := (Id - \Delta_3)$ *defined in* (85), *with homogeneous Neumann boundary conditions is applied to each component of adjoint state vector* $\tilde{\Psi}_0 = (\tilde{\mathbf{V}}, \tilde{\Theta})$.

Remark 2 (Regularity in observational term) Strong solutions of the primitive equations posses regularity in $L^2([0, T], \mathcal{H}^2(\Omega))$ (see Theorem 1.1) but the observational term of the cost functional (70) imposes only the weaker $L^2([0, T], \mathcal{H}^1(\Omega))$-norm. The reason for this discrepancy is that the adjoint equation (54) that is used in Theorem 4.4 requires a forcing $\tilde{F}$ in $L^2([0, T], \mathcal{L}^2(\Omega))$. The adjoint forcing in (91) applies through the operator $\mathcal{S}$ second-order derivatives to the model-observation difference. Taking into account the $L^2([0, T], \mathcal{H}^2(\Omega))$-regularity of the model solution and the imposed regularity of the observations ψ_{obs} the resulting forcing is indeed in $L^2([0, T], \mathcal{L}^2(\Omega))$. The degree of derivatives that appear in $\mathcal{S}$ are linked to the degree of derivatives in the observational part of the cost functional. This becomes evident when the Gateaux-derivative of the cost functional is calculated (see the proof of Theorem 4.4, in particular the integration-by-parts in the second equation of (92)). As this calculation shows, raising the spatial regularity in $\mathcal{J}_{\mathrm{obs}}$ from $\mathcal{H}^1$ to $\mathcal{H}^2$ implies that in the operator $\mathcal{S}$ the derivatives are raised from two to four, such that the highest-order term is then given by the biharmonic operator Δ^2. For such a fourth-order operator, we can no longer guarantee that the adjoint forcing in (91) satisfies $\tilde{F} \in L^2([0, T], \mathcal{L}^2(\Omega))$. The final calculation of the optimal initial state $\bar{\psi}_0$ in (90) restores regularity through the smoothing effect of the inverse $\mathcal{S}^{-1}$ but this requires a square integrable forcing in the adjoint equation.

Proof For a minimizer $\bar\psi_0 := (\bar{\mathbf{v}}_0, \bar\theta_0) \in \mathcal{V}$, which exists according to Theorem 4.1, the Gateaux derivative vanishes such that $\mathcal{J}'(\bar\psi_0; \mathbf{a}) = 0$ for all perturbations $\mathbf{a} = (a_v, a_\theta) \in \mathcal{V}$. We calculate the Gateaux derivative of $\mathcal{J}$ at an arbitrary state ψ in direction $\mathbf{a}$ and with integration by parts follows

$$
\begin{aligned}
\mathcal{J}'(\psi_0; \mathbf{a}) &= \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathbf{a}\rangle_{\mathcal{H}^1} + \int_0^T \int_\Omega \mathcal{R}\left(\mathcal{M}[\psi_0] - \psi_{\text{obs}}\right) \cdot \left(\frac{D\mathcal{M}[\psi_0]}{D\psi_0}\mathbf{a}\right) \, \mathrm{d}x\mathrm{d}t \\
&\quad + \int_0^T \int_\Omega \nabla_3 \mathcal{R}\left(\mathcal{M}[\psi_0] - \psi_{\text{obs}}\right) \cdot \nabla_3 \left(\frac{D\mathcal{M}[\psi_0]}{D\psi_0}\mathbf{a}\right) \, \mathrm{d}x\mathrm{d}t \\
&= \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathbf{a}\rangle_{\mathcal{H}^1} + \int_0^T \int_\Omega \mathcal{R}\left(\mathcal{M}[\psi_0] - \psi_{\text{obs}}\right) \cdot \Psi \, \mathrm{d}x\mathrm{d}t \\
&\quad - \int_0^T \int_\Omega \triangle_3 \mathcal{R}\left(\mathcal{M}[\psi_0] - \psi_{\text{obs}}\right) \cdot \Psi \, \mathrm{d}x\mathrm{d}t \\
&= \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathbf{a}\rangle_{\mathcal{H}^1} + \int_0^T \int_\Omega \mathcal{S}\mathcal{R}\left(\mathcal{M}[\psi_0] - \psi_{\text{obs}}\right) \cdot \Psi \, \mathrm{d}x\mathrm{d}t, \qquad (92)
\end{aligned}
$$

where we have used that $\Psi := \frac{D\mathcal{M}[\psi_0]}{D\psi_0}\mathbf{a}$, according to Lemma 3.2, satisfies the linearized equation with forcing $F = 0$ and that the boundary integral vanishes as a consequence of the boundary conditions.

Now, we integrate an adjoint equation whose forcing is given by the model-observation difference in the second term on the right-hand side of (92),

$$
\tilde{F} := \mathcal{S}\mathcal{R}\left(\mathcal{M}[\psi_0] - \psi_{\text{obs}}\right). \qquad (93)
$$

The initial condition of the adjoint equation is $\tilde\Psi(T) = 0$. From (92) follows with the adjoint relation in Lemma 3.4 that

$$
\begin{aligned}
\mathcal{J}'(\psi_0; \mathbf{a}) &= \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathbf{a}\rangle_{\mathcal{H}^1} + \int_0^T \int_\Omega \tilde{F} \cdot \Psi \, \mathrm{d}x\mathrm{d}t \\
&= \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathbf{a}\rangle_{\mathcal{H}^1} + \int_0^T \int_\Omega F \cdot \tilde\Psi \, \mathrm{d}x\mathrm{d}t - \int_\Omega \tilde\Psi\Psi\big|_0^T \, \mathrm{d}x \\
&= \langle \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathbf{a}\rangle_{\mathcal{H}^1} + \int_\Omega \tilde\Psi(0) \cdot \Psi(0) \, \mathrm{d}x \\
&= \sum_{|\alpha|\leq 1} \langle \mathcal{D}^\alpha \mathcal{B}(\psi_0 - \psi_{\text{back}}), \mathcal{D}^\alpha \mathbf{a}\rangle_{\mathcal{L}^2} + \int_\Omega \tilde\Psi(0) \cdot \mathbf{a} \, \mathrm{d}x \\
&= \int_\Omega \left(\mathcal{S}\mathcal{B}(\psi_0 - \psi_{\text{back}}) + \tilde\Psi(0)\right) \cdot \mathbf{a} \, \mathrm{d}x. \qquad (94)
\end{aligned}
$$

For a minimum we have $\mathcal{J}'(\bar\psi_0; \mathbf{a}) = 0$ for all $\mathbf{a}$, and with (94) we obtain

$$
\mathcal{S}\mathcal{B}(\bar\psi_0 - \psi_{\text{back}}) + \tilde\Psi_0 = 0. \qquad (95)
$$

According to Lemma 4.3 this equations can be solved for $\bar{\psi}_0 = (\bar{\mathbf{v}}_0, \bar{\theta}_0)$ by inverting the operator $\mathcal{S} = (\mathcal{S}_{\mathbf{v}}, \mathcal{S}_{\theta})$. Since $\tilde{\Psi}_0 \in \mathcal{L}^2(\Omega)$, we obtain $\mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}_0 \in \mathcal{H}^2(\Omega)$. This implies with $\psi_{\text{back}} \in \mathcal{H}^1(\Omega)$ that $\bar{\psi}_0 \in \mathcal{H}^1(\Omega)$ is given by

$$\bar{\psi}_0 = \psi_{\text{back}} - \mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}_0.$$

$\square$

We consider now the case of observations that are only square integrable. Consequentially, the H^1-norm in the observational part of the cost functional (70) has to be replaced by the L^2-norm. This results in a different adjoint forcing term (see also remark 2 in the next section). The background term $\mathcal{J}_b$ in (69) remains unchanged, because the H^1-norm is indispensable in obtaining strong solvability of the data assimilation problem.

Corollary 4.5 (Less Regular Observations) *Let observations $\psi_{\text{obs}} \in L^2([0, T], \mathcal{L}^2(\Omega))$ be given. Define the observational term of the cost functional by*

$$\mathcal{J}_{\text{obs}}^*(\psi_0) := \int_0^T \int_\Omega |\mathcal{M}[\psi_0](t) - \psi_{\text{obs}}(t)|^2 \, \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t. \tag{96}$$

Then there exist optimal initial conditions $\bar{\psi}_0 = (\mathbf{v}_0, \theta_0) \in \mathcal{H}^1(\Omega)$ for the data assimilation problem (9) using the cost functional (68) with $\mathcal{J}_{\text{obs}}^*$ replacing $\mathcal{J}_{\text{obs}}$. The minimizer $\bar{\psi}_0$ satisfies

$$\bar{\psi}_0 = \psi_{\text{back}} - \mathcal{S}\mathcal{B}^{-1}\tilde{\Psi}_0, \tag{97}$$

where $\tilde{\Psi}_0$ is the result of integrating the adjoint equation (54) with an initial value $\tilde{\Psi}(T) = 0$ and with the forcing $\tilde{F} := \mathcal{R}(\bar{\psi} - \psi_{\text{obs}})$.

5 Convergence of Gradient-Based Descent Algorithm

The goal of this section is to prove the local convergence of an iterative gradient-based method for determining the optimal initial condition of the data assimilation problem, i.e., the descent method converged provided one uses a starting point that is sufficiently close to the optimal initial condition. Globally convergent algorithms such as the Newton method come with a prohibitively high computational costs for the large-scale problems of numerical weather prediction and ocean state estimation and are not considered here. To prove convergence, we use a general condition that is valid in Hilbert spaces and that relies on the Hessian of the cost functional. The Hessian is calculated through the relation between the Hessian and second-order adjoint equations. This lemma provides the basis of our convergence result.

Lemma 5.1 (Abergel and Temam 1990) *Let J be a real-valued function on a Hilbert space X with norm $|\cdot|$. We make the following assumptions:*

Let $\bar{\psi}$ be an optimal initial condition and denote by $B(\bar{\psi}) \subseteq \mathcal{H}^1(\Omega)$ a ball in $\mathcal{H}^1$ with center $\bar{\psi}$.

Let $\psi_0^0 \in B(\bar{\psi})$ be a start value for the initial condition that lies within a ball around an optimal initial condition. Choose $\lambda_0 > 0$ and iterate for $n = 1 \ldots$

1. *Integrate the primitive equations (1) with initial condition ψ_0^n to compute the model trajectory $\psi^n(t), t \in [0, T]$.*

2. *To calculate the adjoint state $\tilde{\Psi}^n(t)$, integrate the adjoint equation (39) over $[0, T]$ with initial condition $\tilde{\Psi}^n(T) = 0$ and forcing $\tilde{F} := \mathcal{SR}(\psi^n - \psi_{obs})$.*

3. *Apply to the final adjoint state $\tilde{\Psi}_0^n$ the smoothing operator $\mathcal{S}^{-1}$, defined in Theorem 4.4, to compute $\mathcal{S}^{-1}\tilde{\Psi}_0^n$.*

4. *Define the update ψ_0^{n+1} to initial velocity by*

$$\psi_0^{n+1} := \psi_0^n + \lambda_n \mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}^n. \tag{98}$$

5. *Increase n, update the stepsize λ_n such that*

$$\mathcal{J}(\psi_0^n + \lambda_n \mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}^n) = \inf \big\{ \mathcal{J}(\psi_0^n + \lambda \mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}^n) : \lambda > 0$$
$$\text{such that } \psi_0^n + \lambda \mathcal{B}^{-1}\mathcal{S}^{-1}\tilde{\Psi}^n \in B \big\}. \tag{99}$$

Fig. 1 Gradient algorithm for calculation of minimizers of cost functional. The difference between the $\mathcal{H}^1$-norms and the $\mathcal{L}^2$-norms in the cost functional appear in steps 2 and 3. For the $\mathcal{L}^2$ version, the operator $\mathcal{S}$ in the adjont forcing $\tilde{F}$ is the identity and step 3 does not exist

(i) *J is of class C^2 and has a local minimum at a point $\bar{x} \in X$,*
(ii) *there exists a ball $B(\bar{x}) \subseteq X$ around $\bar{x}$, and two real numbers m, M, such that the following inequalities hold:*

$$m|x||y| \leq J''(u; x, y) \leq M|x||y|, \quad \text{for all } u \in B, \text{ and } x, y \in X,$$

where $J''[u; x, y]$ is the bilinear form associated with the second derivative of J. Then, the gradient algorithm with initial value $x_0 \in B$ converges to $\bar{x}$. The gradient algorithm is shown in Figure 1 where the cost functional J corresponds to $\mathcal{J}$ in Fig. 1, the local minimum $\bar{x}$ to $\bar{\psi}$ and the initial value x_0 to ψ_0^0.

The adjoint gradient-based algorithm for which we want to establish convergence is illustrated in Fig. 1.

The second derivative of the cost functional $\mathcal{J}$ is related to the Hessian $H_{\mathcal{J}}[\psi]$ via

$$\mathcal{J}''(\psi_0; \mathcal{W}, \mathcal{Z}) = \langle \mathcal{Z}, H_{\mathcal{J}}[\psi]\mathcal{W} \rangle, \quad \text{for } \mathcal{W}, \mathcal{Z} \in H^1(\Omega). \tag{100}$$

1. *Integration of the primitive equations (1) over $[0, T]$ with initial condition ψ_0 to obtain $\psi(t)$.*

2. *Integration of the linearized model (32) with initial condition $\Psi_0 = \mathcal{W}$, vanishing forcing, and nonlinear state $\psi(t)$.*

3. *Integration of the adjoint model (54) with initial condition $\tilde{\Psi}(T) = 0$ and forcing $\tilde{F} := \sum_{|\alpha| \leq 1} \triangle^{\alpha} \mathcal{R}(\psi - \psi_{obs})$.*

4. *Integration of the second order adjoint (101) with initial condition $\hat{\Psi}(T) = 0$ and forcing $\bar{F} = \sum_{|\alpha| \leq 1} \triangle^{\alpha} \mathcal{R} \Psi$.*

5. *The final state of the second order adjoint at $t = 0$, $\hat{\Psi}_0$ satisfies*

$$H_{\mathcal{J}}[\psi]\mathcal{W} = \mathcal{SBW} - \hat{\Psi}(t_0) \qquad \text{for } \mathcal{W} \in \mathcal{H}^1(\Omega). \tag{103}$$

Fig. 2 Algorithm for calculation Hesse matrix via second-order adjoint

The calculation of the Hessian $H_{\mathcal{J}}$ of the cost functional uses the second-order adjoint equations. The second-order adjoint equations are given by (for details see Appendix 1)

$$
\begin{aligned}
&- \partial_t \hat{U} + u \partial_x \hat{U} + \hat{U} \partial_x u + v \partial_y \hat{U} + w \partial_z \hat{U} + \hat{V} \partial_y u - f \hat{V} + \partial_x \\
&\int_{-h}^{z} g \hat{\Theta}(x, y, \xi, t)\, d\xi + D_v(\hat{U}) = \bar{F}_{\hat{U}} + \mathcal{G}_{\hat{U}}, \\
&- \partial_t \hat{V} + \hat{U} \partial_x v + v \partial_y \hat{V} + \hat{V} \partial_y v + u \partial_x \hat{V} + w \partial_z \hat{V} + f \hat{U} + \partial_y \\
&\int_{-h}^{z} g \hat{\Theta}(x, y, \xi, t)\, d\xi + D_v(V) = \bar{F}_{\hat{V}} + \mathcal{G}_{\hat{V}}, \\
&\partial_x \hat{U} + \partial_y \hat{V} + \partial_z \bar{W} = 0, \\
&- \partial_t \hat{\Theta} + u \partial_x \hat{\Theta} + v \partial_y \hat{\Theta} + w \partial_z \hat{\Theta} + \hat{U} \partial_x \theta + \hat{V} \partial_y \theta + D_{\theta}(\hat{\Theta}) = \bar{F}_{\hat{\Theta}} + \mathcal{G}_{\hat{\Theta}},
\end{aligned}
\tag{101}
$$

where $D_v(f) := -\frac{1}{Re_1} \triangle f - \frac{1}{Re_2} \partial_{zz}^2 f$ and $D_{\theta}(\theta) := -\frac{1}{Rt_1} \triangle \theta - \frac{1}{Rt_2} \partial_{zz}^2 \theta$ and with forcing $\bar{F} := (\bar{F}_{\hat{U}}, \bar{F}_{\hat{V}}, \bar{F}_{\hat{\Theta}})$ and with $\mathcal{G} := (\mathcal{G}_{\hat{U}}, \mathcal{G}_{\hat{V}}, \mathcal{G}_{\hat{\Theta}})$ given by

$$
\begin{aligned}
\mathcal{G}_{\hat{U}} &:= U \partial_x \tilde{U} + \tilde{U} \partial_x U + V \partial_y \tilde{U} + \tilde{U} \partial_x V + \tilde{U} \partial_x \Theta + \tilde{\Theta} \partial_x \Theta, \\
\mathcal{G}_{\hat{V}} &:= \tilde{V} \partial_y U + V \partial_y \tilde{V} + \tilde{V} \partial_y + U \partial_x \tilde{V} + \tilde{V} \partial_y \Theta + \tilde{\Theta} \partial_y \Theta, \\
\mathcal{G}_{\hat{\Theta}} &:= \partial_x (U \tilde{\Theta}) + \partial_y (V \tilde{\Theta}) + \partial_z (W \tilde{\Theta}) + \int_{-h}^{z} \nabla \cdot \tilde{U}(x, y, z')\, dz'.
\end{aligned}
\tag{102}
$$

The following Theorem provides information about the regularity of the second-order adjoint equations that we need to prove our convergence result.

Theorem 5.2 (Regularity of Second-Order Adjoint Equations) *Suppose*

1. *the initial condition $\psi_0 \in \mathcal{V}$ of the primitive equations* (1) *is given*
2. *the initial condition $\Psi_0 \in \mathcal{V}$ of the linearized equations* (32) *is given*
3. *the initial condition $\tilde{\Psi}(T) \in \mathcal{V}$ of the adjoint equations* (54)*, specified at the end point of the time interval is given*
4. *the forcing satisfies $\hat{F} = (\hat{F}_{\hat{U}}, \hat{F}_{\hat{V}}, \hat{F}_{\hat{\Theta}}) \in L^2([0, T], \mathbf{L}^2(\Omega) \times L^2(\Omega))$.*

If the initial conditions of the second-order adjoint equations satisfy $\hat{\Psi}_0 \in \mathcal{V}$, then the system (101) *has a unique solution on $[0, T]$ with the properties*

$$\hat{\Psi}(t) \in C([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega)),$$

and the state vector $\hat{\Psi}$ satisfies for $t \in [0, T]$

$$
\begin{aligned}
||\hat{\Psi}(t)||^2_{\mathcal{H}^1} &\leq ||\hat{\Psi}_0||^2_{\mathcal{H}^1} \exp\left\{ \int_0^t K_1(y) + K_2(y) \, \mathrm{d}y \right\} \\
&+ \int_0^t \left[||\hat{F}||^2_{\mathcal{L}^2} + ||\mathcal{G}||^2_{\mathcal{L}^2} \right] \exp\left\{ \int_0^s K_1(y) + K_2(y) \, \mathrm{d}y \right\} \mathrm{d}s,
\end{aligned}
\tag{104}
$$

with K_1, K_2 defined in (44) *and $\mathcal{G}$ defined in* (102)*.*

Proof *(Sketch of proof)* The second-order adjoint equations (101) resemble formally the first-order adjoint equations (54). If one identifies the first-order adjoint variable $\tilde{\Psi}$ with the second-order adjoint variable $\bar{\Psi}$, then the difference between the two equations are the additional $\mathcal{G}$-terms defined in (102). These terms consist of products of linear variable $\Psi := (\mathbf{V}, \Theta)$ and derivatives of the second order adjoint $\bar{\Psi}$. Using the regularity of linear equations in Theorem 3.1, we can estimate $\mathcal{G}_{\hat{U}}, \mathcal{G}_{\hat{V}}, \mathcal{G}_{\hat{\Theta}}$ in the same manner as the products of the state vector ψ and derivatives of the second-order adjoint $\bar{\Psi}$ on the left-hand side of (101). With the Agmon inequality, we obtain the following estimate

$$
\begin{aligned}
\int_0^T ||\mathcal{G}_{\hat{U}}(s)||^2_{\mathcal{L}^2} \mathrm{d}s &\leq \max_{s \in [0,T]} ||\nabla \widetilde{\mathbf{V}}(s)||^2_{\mathbf{L}^2} \int_0^T ||\mathbf{V}(s)||^2_{\mathbf{H}^2} \mathrm{d}s \\
&+ \max_{s \in [0,T]} ||\nabla \mathbf{V}(s)||^2_{\mathbf{L}^2} \int_0^T ||\widetilde{\mathbf{V}}(s)||^2_{\mathbf{H}^2} \mathrm{d}s \\
&+ \max_{s \in [0,T]} ||\nabla \Theta(s)||^2_{L^2} \int_0^T ||\widetilde{\mathbf{V}}(s)||^2_{\mathbf{H}^2} + ||\widetilde{\Theta}(s)||^2_{\mathbf{H}^2} \mathrm{d}s.
\end{aligned}
\tag{105}
$$

Analogous estimates for $\mathcal{G}_{\hat{V}}, \mathcal{G}_{\hat{\Theta}}$ imply that $\mathcal{G} = (\mathcal{G}_{\hat{U}}, \mathcal{G}_{\hat{V}}, \mathcal{G}_{\hat{\Theta}})$ satisfies

$$
\begin{aligned}
\int_0^T ||\mathcal{G}(s)||^2_{\mathcal{L}^2} \mathrm{d}s &\leq ||\Psi(s)||^2_{C([0,T],\mathcal{H}^1)} ||\widetilde{\Psi}(s)||^2_{L^2([0,T],\mathcal{H}^2)} \\
&+ ||\widetilde{\Psi}(s)||^2_{C([0,T],\mathcal{H}^1)} ||\Psi(s)||^2_{L^2([0,T],\mathcal{H}^2)}.
\end{aligned}
\tag{106}
$$

This shows that $(\mathcal{G}_{\hat{U}}, \mathcal{G}_{\hat{V}}, \mathcal{G}_{\hat{\Theta}}) \in L^2([0, T], \mathcal{L}^2)$. If we now define $\tilde{F} := \mathcal{G} + \bar{F}$, we can cast the second-order adjoint equations in the form of the first-order adjoint equations and apply the estimates in the proof of Theorem 3.3 to prove that a unique solution $\bar{\Psi} \in C([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega))$ to (101) exists.　　　　□

The equations (100) and (103) are used to verify the boundedness of the second derivative of the cost functional. From Theorem 5.2, we infer that the right-hand side of (103) is well-defined in $\mathcal{H}^1(\Omega)$, i.e., for $\mathcal{W} \in \mathcal{H}^1(\Omega)$, we have

$$H_{\mathcal{J}}[\psi]\mathcal{W} = \mathcal{SBW} - \bar{\Psi}_0. \tag{107}$$

We show now that the algorithm described in Figure 1 converges to an optimal initial condition of the data assimilation problem.

Theorem 5.3 (Convergence) *Suppose that the assumptions of Theorem 4.1 are satisfied. Let $\bar{\psi}_0 \in \mathcal{V}$ be an optimal initial condition for the data assimilation problem (9) with cost functional specified by (68). Let $\psi_0^0 \in \mathcal{V}$ be an initial value for the descent algorithm 5 that lies within a ball $B(\bar{\psi}_0) \subseteq \mathcal{H}^1(\Omega)$ around $\bar{\psi}_0$. Define the sequence $(\psi_0^n)_n$ by (98). Then, $(\psi_0^n)_n$ converges to $\bar{\psi}_0$ in $\mathcal{H}^1(\Omega)$.*

Proof The proof of the Theorem is based on Lemma 5.1. We have to establish a lower and upper bound on the Hessian of the cost functional. This is accomplished via bounds on the second-order state variable. In order to ease notation we suppress the index "n" of the sequence, the state variable ψ in this proof corresponds to ψ_n. The occurring equations, their initial conditions and forcing terms are summarized in Fig. 2.

We infer from Theorem 5.2 with $\hat{\Psi}_0 = 0$

$$||\hat{\Psi}(t)||^2_{\mathcal{H}^1} \le \int_t^T \left[||\hat{F}||^2_{\mathcal{L}^2} + ||\mathcal{G}||^2_{\mathcal{L}^2} \right] \exp \left\{ \int_0^s K_1(y) + K_2(y)\, \mathrm{d}y \right\} \mathrm{d}s, \tag{108}$$

with K_1, K_2 defined in (44) and with forcing given by $\hat{F} := \sum_{|\alpha| \le 1} \triangle^\alpha \mathcal{R}\Psi$.

We estimate now the right-hand side of (108) and begin with the forcing. With assumption (71) on the covariance operators, we obtain the estimate

$$\int_t^T ||\hat{F}(s)||^2_{\mathcal{L}^2}\mathrm{d}s \le \int_t^T \sum_{|\alpha| \le 1} ||\triangle^\alpha \mathcal{R}\Psi(s)||^2_{\mathcal{L}^2}\mathrm{d}s \le c \int_t^T ||\mathcal{R}\Psi(t)||^2_{\mathcal{H}^2}\mathrm{d}s$$

$$\le c||\mathcal{R}|| \int_t^T ||\Psi(t)||^2_{\mathcal{H}^2}\mathrm{d}s, \tag{109}$$

where $||\mathcal{R}||$ denotes the operator norm of the linear operator $\mathcal{R}$ and $\Psi \in L^\infty([0, T], \mathcal{H}^1(\Omega)) \cap L^2([0, T], \mathcal{H}^2(\Omega))$ is the solution of the linear primitive equations with initial condition $\Psi_0 = \mathcal{W}$ and vanishing forcing. According to (42), this solution satisfies for $t \in [0, T]$ the inequalities

$$||\Psi(t)||^2_{\mathcal{H}^1} \le ||\mathcal{W}||^2_{\mathcal{H}^1} \exp\{ \int_0^t K_1(s) + K_2(s)\, \mathrm{d}s \} = L_0(t)||\mathcal{W}||^2_{\mathcal{H}^1},$$

$$\text{with} \quad L_0(t) := \exp\{\int_0^t K_1(s) + K_2(s)\,ds\}, \tag{110}$$

and

$$\int_t^T ||\Psi(s)||_{\mathcal{H}^2}^2 ds$$

$$\leq ||\mathcal{W}||_{\mathcal{H}^1}^2 + ||\mathcal{W}||_{\mathcal{H}^1}^2 \int_t^T (K_1(s) + K_2(s)) \exp\{\int_0^s K_1(y) + K_2(y)\,dy\}\,ds$$

$$\leq ||\mathcal{W}||_{\mathcal{H}^1}^2 \left(1 + \int_t^T (K_1(s) + K_2(s)) \exp\{\int_0^s K_1(y) + K_2(y)\,dy\}\,ds\right)$$

$$= L_1(t)||\mathcal{W}||_{\mathcal{H}^1}^2,$$

$$\text{with} \quad L_1(t) := \left(1 + \int_t^T (K_1(s) + K_2(s)) \exp\{\int_0^s K_1(y) + K_2(y)\,dy\}\,ds\right). \tag{111}$$

The estimates (110) and (111) imply for (109)

$$\int_t^T ||\hat{F}(s)||_{\mathcal{L}^2}^2 ds \leq M_1(t)||\mathcal{W}||_{\mathcal{H}^1}^2, \tag{112}$$

$$\text{with } M_1(t) := c(L_0(t) + L_1(t))||\mathcal{R}||.$$

We continue now with the estimation of the $\mathcal{G}$-terms in (108). It follows analogous to (106)

$$\int_t^T ||\mathcal{G}(s)||_{\mathcal{L}^2}^2 ds \leq ||\Psi||_{C([t,T],\mathcal{H}^1)}^2 ||\widetilde{\Psi}||_{L^2([t,T],\mathcal{H}^2)}^2 \tag{113}$$

$$+ ||\widetilde{\Psi}||_{C([t,T],\mathcal{H}^1)}^2 ||\Psi||_{L^2([t,T],\mathcal{H}^2)}^2.$$

The adjoint state $\widetilde{\Psi}$ in (113) which result from integration of the adjoint equations with zero terminal condition and forcing $\tilde{F} := \sum_{|\alpha|\leq 1} \Delta^\alpha \mathcal{R}(\psi - \psi_{\text{obs}})$ satisfies according to Theorem 3.3 for $t \in [0, T]$ the H^1-estimate

$$||\widetilde{\Psi}(t)||_{\mathcal{H}^1}^2 \leq c \int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2 \exp\{\int_0^s K_1(y) + K_2(y)\,dy\}\,ds \tag{114}$$

$$\leq c||L_0||_{L^\infty} \int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2 \,ds.$$

For the H^2-bound on the adjoint state in (113) yields Theorem 3.3, combined with (114)

$$
\begin{aligned}
\int_t^T ||\widetilde{\Psi}(s)||_{\mathcal{H}^2}^2 ds &\leq \int_t^T (K_1(s) + K_2(s))||\widetilde{\Psi}(s)||_{\mathcal{H}^1}^2 \, ds + \int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2 \, ds \\
&\leq ||L_0||_{L^\infty} \int_t^T (K_1(s) + K_2(s)) \int_0^s ||\tilde{F}(y)||_{\mathcal{L}^2}^2 \, dy \, ds + \int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2, \, ds \\
&\leq ||L_0||_{L^\infty} \int_t^T (K_1(s) + K_2(s)) \, ds \int_0^t ||\tilde{F}(y)||_{\mathcal{L}^2}^2 \, dy \\
&\quad + \int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2, \, ds \\
&\leq L_2(t) \int_t^T ||\tilde{F}(y)||_{L^2}^2 \, dy,
\end{aligned}
\tag{115}
$$

with $L_2(t) := ||L_0||_{L^\infty} \int_t^T (K_1(s) + K_2(s)) \, ds + 1$.

For the adjoint forcing in (114) and (115) follows with the triangle inequality

$$
\begin{aligned}
\int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2 ds &\leq \sum_{|\alpha| \leq 1} \int_0^t ||\Delta^\alpha \mathcal{R}\left(\psi(s) - \psi_{\text{obs}}(s)\right)||_{\mathcal{L}^2}^2 ds \\
&\leq ||\mathcal{R}|| \int_0^t ||\psi(s) - \psi_{\text{obs}}(s)||_{\mathcal{H}^2}^2 ds \\
&\leq ||\mathcal{R}|| \int_t^T ||\psi(s) - \bar{\psi}(s)||_{\mathcal{H}^2}^2 + ||\bar{\psi}(s) - \psi_{\text{obs}}(s)||_{\mathcal{H}^2}^2 ds =: L_4(t),
\end{aligned}
\tag{116}
$$

where $L_4(t) > 0$ is bounded on T, since $\psi^*, \psi^n, \psi_{\text{obs}} \in L^2([0, T], \mathcal{H}^2(\Omega))$. The function L_4 is bounded uniformly in n, because $(\psi^n)_n \subseteq B(\psi_0^*)$.

From (114)-(116) follows for (113)

$$
\begin{aligned}
\int_t^T ||\mathcal{G}(s)||_{\mathcal{L}^2}^2 ds & \\
&\leq ||\Psi||_{C([t,T]H^1)}^2 ||\widetilde{\Psi}||_{L^2([t,T],\mathcal{H}^2)}^2 + ||\widetilde{\Psi}||_{C([t,T]H^1)}^2 ||\Psi||_{L^2([t,T],\mathcal{H}^2)}^2 \\
&\leq (L_0(t)L_2(t) + c||L_0||_{L^\infty} L_1(t)) \, ||\mathcal{W}||_{\mathcal{H}^1}^2 \int_t^T ||\tilde{F}(s)||_{\mathcal{L}^2}^2 \, ds \\
&\leq (L_0(t)L_2(t) + c||L_0||_{L^\infty} L_1(t)) \, L_4(t)||\mathcal{W}||_{H^1} = M_2(t)||\mathcal{W}||_{\mathcal{H}^1}^2,
\end{aligned}
\tag{117}
$$

with $M_2(t) := (L_0(t)L_2(t) + c||L_0||_{L^\infty} L_1(t)) \, L_4(t)$

With (109) and (117), we derive for the upper bound on the second-order state in (108) for $t \in [0, T]$

$$
\begin{aligned}
||\hat{\Psi}(t)||_{\mathcal{H}^1}^2 &\leq \int_t^T \left[||\hat{F}||_{\mathcal{L}^2}^2 + ||\mathcal{G}||_{\mathcal{L}^2}^2 \right] \exp\{ \int_0^s K_1(y) + K_2(y) \, \mathrm{d}y \} \, \mathrm{d}s \\
&\leq L_0(t)(M_1(t) + M_2(t))||\mathcal{W}||_{\mathcal{H}^1} = M_0(t))||\mathcal{W}||_{\mathcal{H}^1}, \\
&\text{with } M_0(t) := L_0(t)(M_1(t) + M_2(t)).
\end{aligned}
\tag{118}
$$

We are now in the position to derive upper and lower bounds on the Hessian. From (118) follows

$$
\begin{aligned}
|\langle \mathcal{Z}, H_{\mathcal{J}}[\psi]\mathcal{W}\rangle_{\mathcal{H}^1}| &= |\langle \mathcal{Z}, \mathcal{SBW} - \hat{\Psi}_0\rangle_{\mathcal{H}^1}| \\
&\leq ||\mathcal{Z}||_{\mathcal{H}^1} ||\mathcal{SBW} - \hat{\Psi}_0||_{\mathcal{H}^1} \\
&\leq ||\mathcal{Z}||_{\mathcal{H}^1} (||\mathcal{SBW}||_{\mathcal{H}^1} + ||\hat{\Psi}_0||_{\mathcal{H}^1}) \\
&\leq ||\mathcal{Z}||_{\mathcal{H}^1} (||\mathcal{SB}||_{\mathcal{H}^1} + \sqrt{M_0}|)\mathcal{W}||_{\mathcal{H}^1}.
\end{aligned}
\tag{119}
$$

Equations (119) and (100) establish the bounds on the second derivative of the cost functional and the application of Lemma 5.1 proves the assertion of the Theorem. □

Appendix A: Derivation of First Order Linear, Adjoint and Second Order Adjoint Primitive Equations

Let $\psi := (u, v, \theta)$ be the state vector of the primitive equations. We write the primitive equations (1) in the form

$$
\frac{\mathrm{d}}{\mathrm{d}t}\psi = F(\psi),
\tag{120}
$$

with

$$
F(\psi) = \begin{pmatrix} u\partial_x u + v\partial_y u + w\partial_z u - fv + \partial_x \int_{-h}^z g\rho(x, y, \xi, t) \, \mathrm{d}\xi + D(u) \\ v\partial_x v + v\partial_y v + w\partial_z v + fu + \partial_y \int_{-h}^z g\rho(x, y, \xi, t) \, \mathrm{d}\xi + D(v) \\ u\partial_x\theta + v\partial_y\theta + w\partial_z\theta + D_\theta(\theta), \end{pmatrix}.
\tag{121}
$$

where $D_v(f) := -\frac{1}{Re_1} \triangle f - \frac{1}{Re_2} \partial_{zz}^2 f$ and $D_\theta(\theta) := -\frac{1}{Rt_1} \triangle \theta - \frac{1}{Rt_2} \partial_{zz}^2 \theta$.

Consider now a linearized perturbation $\Psi := (U, V, \Theta)$ of the state vector $\psi = (u, v, \theta)$ that results from a perturbation of the initial condition $\psi_0 := (u_0, v_0, \theta_0)$ of the state vector. Note that we only perturb variables that require an initial condition, i.e., for the primitive equations and in contrast to the Navier–Stokes equations we do not perturb the vertical velocity. The perturbation of u for example is defined as Gateaux derivative of u at u_0 in direction h

$$U := \frac{d}{dt} u(u_0 + th)|_{t=0}, \tag{122}$$

and analogously for v and θ. The equation governing the dynamics of the linearized perturbation Ψ can be found by taking the Gateaux derivative of the primitive equations (1). The product of the Jacobi-matrix F', consisting of the partial Gateaux derivatives of F with respect to the components (u, v, θ) of the state vector ψ, applied to the linearized state vector $\Psi = (U, V, \Theta)$ is given by

$$F'[\psi](\Psi) =$$
$$\begin{pmatrix} u\partial_x U + U\partial_x u + v\partial_y U + V\partial_y u + w\partial_z U + W\partial_z u & -fV + \partial_x \int_{-h}^z g\Theta(x, y, \xi, t)\, d\xi & +D_v(U) \\ U\partial_x v + v\partial_y V + V\partial_y v + u\partial_x V + w\partial_z V + W\partial_z v & +fU + \partial_y \int_{-h}^z g\Theta(x, y, \xi, t)\, d\xi & +D_v(V) \\ \partial_x(u\Theta) + \partial_y(v\Theta) + \partial_z(w\Theta) + \partial_x(U\theta) + \partial_y(V\theta) + \partial_z(W\theta) + D_\theta(\Theta) \end{pmatrix}.$$
$$\tag{123}$$

The linearized primitive equations can be written in the form

$$\frac{d}{dt} \Psi = F'[\psi](\Psi). \tag{124}$$

subject to the divergence-free constraint $\nabla_3 \cdot \mathbf{V} = 0$.

The adjoint primitive equation with respect to the L^2-scalar product reads as follows

$$-\frac{d}{dt} \tilde{\Psi} = F'^*[\psi](\tilde{\Psi}). \tag{125}$$

The expression of the adjoint operator $F'^*[\psi]$ is derived by taking L^2-scalar product of the linear operator $F'[\psi]$ with the adjoint state vector $\tilde{\Psi} = (\tilde{\mathbf{V}}, \tilde{\Theta})$, $\tilde{\mathbf{V}} := (\tilde{U}, \tilde{V})$ performing integration-by-parts and rearranging the resulting terms such that the linear state vector Ψ takes the role of a test function and the other component of the scalar product consists of adjoint variables $\tilde{\Psi}$ and state variable ψ. We illustrate this with the pressure gradient in the velocity equation. Integrating by parts, using the homogeneous Dirichlet boundary conditions and the incompressibility of the adjoint velocity field yields

$$\int_\Omega \nabla \left(\int_{-h}^z \Theta(x, y, \xi)\, d\xi \right) \cdot \tilde{\mathbf{V}}(x, y, z)\, dx\,dy\,dz$$

$$
= -\int_{\Omega} \left(\int_{-h}^{z} \Theta(x, y, \xi)\, d\xi \right) \nabla \cdot \tilde{\mathbf{V}}(x, y, z)\, dxdydz
$$

$$
+ \int_{\partial\Omega} \left(\int_{-h}^{z} \Theta(x, y, \xi)\, d\xi \right) \vec{n}_H \cdot \tilde{\mathbf{V}}(x, y)\, ds
$$

$$
= \int_{\Omega} \left(\int_{-h}^{z} \Theta(x, y, \xi)\, d\xi \right) \partial_z \tilde{W}(x, y, z)\, dxdydz
$$

$$
= -\int_{\Omega} \Theta(x, y, z) \tilde{W}(x, y, z)\, dxdydz + \left[\int_{-h}^{z} \Theta(x, y, \xi)\, d\xi\ \tilde{W}(x, y, z) \right]_{z=-h}^{z=0}
$$

$$
= \int_{\Omega} \Theta(x, y, z) \left(\int_{-h}^{z} \nabla \cdot \tilde{\mathbf{V}}(x, y, \xi) d\xi \right) dxdydz, \tag{126}
$$

where $\vec{n}_H$ denotes the horizontal outer normal vector. The multiplication by the linear temperature variable Θ as a test function implies that the vertical integral over the horizontal divergence is an element of the weak formulation of the adjoint temperature equation. This results in the following expression for $F'^{*}[\psi]$

$$
F'^{*}[\psi](\tilde{\Psi})
$$
$$
= \begin{pmatrix}
u\partial_x\tilde{U} + \tilde{U}\partial_x u + v\partial_y\tilde{U} + \tilde{U}\partial_x v + w\partial_z\tilde{U} + \tilde{W}\partial_z u + f\tilde{U} + \tilde{U}\partial_x\theta + \theta\partial_x\tilde{\Theta} + D_v(\tilde{U}) \\
\tilde{V}\partial_y u + v\partial_y\tilde{V} + \tilde{V}\partial_y v + u\partial_x\tilde{V} + w\partial_z\tilde{V} + \tilde{W}\partial_z v - f\tilde{V} + \tilde{V}\partial_y\theta + \theta\partial_y\tilde{\Theta} + D_v(\tilde{V}) \\
u\partial_x(\tilde{\Theta}) + v\partial_y(\tilde{\Theta}) + w\partial_z(\tilde{\Theta}) + \int_{-h}^{z} \nabla \cdot \tilde{U}(x, y, z')\, dz' + D_\theta(\tilde{\Theta})
\end{pmatrix},
$$
$$\tag{127}$$

subject to the constraint $\partial_x\tilde{U} + \partial_y\tilde{V} + \partial_z\tilde{W} = 0$.

In order to derive the second-order adjoint primitive equations we linearize the combined set of primitive and adjoint equations. The resulting second-order adjoint primitive equations are given by

$$
F''^{*}[\psi, \Psi](\hat{\Psi})
$$
$$
= \begin{pmatrix}
u\partial_x\hat{U} + \hat{U}\partial_x u + v\partial_y\hat{U} + \hat{V}\partial_y u + w\partial_z\hat{U} + \hat{W}\partial_z u\ -f\hat{V} + \partial_x \int_{-h}^{z} g\hat{\Theta}(x, y, \xi, t)\, d\xi\ + D_v(\hat{U}) \\
\hat{U}\partial_x v + v\partial_y\hat{V} + \hat{V}\partial_y v + u\partial_x\hat{V} + w\partial_z\hat{V} + \hat{W}\partial_z v\ + f\hat{U} + \partial_y \int_{-h}^{z} g\hat{\Theta}(x, y, \xi, t)\, d\xi\ + D_v(V) \\
u\partial_x\hat{\Theta} + v\partial_y\hat{\Theta} + w\partial_z\hat{\Theta} + D_\theta(\hat{\Theta})
\end{pmatrix}
$$
$$
+ \begin{pmatrix}
U\partial_x\tilde{U} + \tilde{U}\partial_x U + V\partial_y\tilde{U} + \tilde{U}\partial_x V & +\tilde{U}\partial_x\Theta + \tilde{\Theta}\partial_x\Theta \\
\tilde{V}\partial_y U + V\partial_y\tilde{V} + \tilde{V}\partial_y + U\partial_x\tilde{V} & +\tilde{V}\partial_y\Theta + \tilde{\Theta}\partial_y\Theta \\
\partial_x(U\tilde{\Theta}) + \partial_y(V\tilde{\Theta}) + \partial_z(W\tilde{\Theta}) & + \int_{-h}^{z} \nabla \cdot \tilde{U}(x, y, z')\, dz'
\end{pmatrix}.
$$
$$\tag{128}$$

Appendix B: Proof of Theorem 3.1

Proof The following proof is formal, it can be easily made rigorous by using a Galerkin expansion of the linearized variables $(\mathbf{V}, \Theta)$. The a priori estimates of the following proof, applied to this Galerkin expansion, allows then to pass to the limit with the Lions–Aubin compactness lemma.

Techniques that were originally developed in Cao and Titi (2007) are instrumental for the proof.

Step 1: $L^\infty([0, T], L^2(\Omega))$-bound on velocity.

Taking the L^2-inner product of the linearized velocity equation (32a) with the velocity $\mathbf{V}$ yields

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}\,\|\mathbf{V}\|_{\mathbf{L}^2}^2}{\mathrm{d}t} &+ \frac{1}{Re_1}\|\nabla\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{1}{Re_2}\|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2 \\
&= -\int_\Omega \left[(\mathbf{v}\cdot\nabla)\mathbf{V} + (\mathbf{V}\cdot\nabla)\mathbf{v} - \left(\int_{-h}^z \nabla\cdot\mathbf{v}(x,y,\xi,t)\,\mathrm{d}\xi\right)\partial_z\mathbf{V} \right]\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad - \left(\int_{-h}^z \nabla\cdot\mathbf{V}(x,y,\xi,t)\,\mathrm{d}\xi\right)\partial_z v\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z - \int_\Omega (\nabla P_s + f\vec{k}\times\mathbf{V})\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad - \int_\Omega \nabla\left(\int_{-h}^z \Pi(x,y,\xi,t)\mathrm{d}\xi\right)\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega \mathbf{V}\cdot F_\mathbf{V}\mathrm{d}x\mathrm{d}y\mathrm{d}z.
\end{aligned}
\tag{129}
$$

The term with the Coriolis force vanishes. For the surface pressure we obtain after integration by parts

$$
-\int_\Omega \nabla P_s\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_\Omega P_s\nabla\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_{\Gamma_u} P_s(\nabla\cdot\int_{-h}^0 \mathbf{V}\,\mathrm{d}z)\,\mathrm{d}x\mathrm{d}y = 0.
$$

For the linearized temperature equation we obtain after taking the L^2-inner product of (32c) with the temperature Θ

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}\|\Theta\|_{L^2}^2}{\mathrm{d}t} &+ \frac{1}{Rt_1}\int_\Omega |\nabla\Theta|^2\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + \frac{1}{Rt_2}\int_\Omega |\partial_z\Theta|^2\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + k_\theta\|\theta(z=0)\|_{L^2}^2 \\
&= -\int_\Omega \left[(\mathbf{v}\cdot\nabla)\Theta + (\mathbf{V}\cdot\nabla)\theta - \left(\int_{-h}^z \nabla\cdot\mathbf{v}(x,y,\xi,t)\,\mathrm{d}\xi\right)\partial_z\Theta \right]\Theta\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad - \int_\Omega \left(\int_{-h}^z \nabla\cdot\mathbf{V}(x,y,\xi,t)\,\mathrm{d}\xi\right)\partial_z\theta\,\Theta\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega \Theta\cdot F_\Theta\mathrm{d}x\mathrm{d}y\mathrm{d}z.
\end{aligned}
\tag{130}
$$

Using that the following advective terms vanish due to the incompressibility (1c) of the velocity **v**

$$\int_\Omega \left[(\mathbf{v} \cdot \nabla)\mathbf{V} - \left(\int_{-h}^z \nabla \cdot \mathbf{v}(x, y, \xi, t)\, d\xi \right) \partial_z \mathbf{V} \right] \cdot \mathbf{V}\, dx dy dz = 0,$$

$$\int_\Omega \left[(\mathbf{v} \cdot \nabla)\Theta - \left(\int_{-h}^z \nabla \cdot \mathbf{v}(x, y, \xi, t)\, d\xi \right) \partial_z \Theta \right] \Theta\, dx dy dz = 0,$$

yields for the velocity equation

$$\frac{1}{2} \frac{d}{dt} ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re_1} ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re_2} ||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2$$
$$= - \int_\Omega \left[(\mathbf{V} \cdot \nabla)\mathbf{v} - \left(\int_{-h}^z \nabla \cdot \mathbf{V}(x, y, \xi, t)\, d\xi \right) \partial_z \mathbf{v} \right] \cdot \mathbf{V}\, dx dy dz \qquad (131)$$
$$+ \int_\Omega \left[\nabla \left(\int_{-h}^z \Pi(x, y, \xi, t) d\xi \right) \right] \cdot \mathbf{V}\, dx dy dz + \int_\Omega \mathbf{V} \cdot F_\mathbf{V}\, dx dy dz.$$

For the temperature equation, we obtain

$$\frac{1}{2} \frac{d}{dt} ||\Theta||_{L^2}^2 + \frac{1}{Rt_1} \int_\Omega |\nabla \Theta|^2\, dx dy dz + \frac{1}{Rt_2} \int_\Omega |\partial_z \Theta|^2\, dx dy dz + k_\theta ||\theta(z = 0)||_{L^2}^2$$
$$= - \int_\Omega \left[(\mathbf{V} \cdot \nabla)\theta - \left(\int_{-h}^z \nabla \cdot \mathbf{V}(x, y, \xi, t)\, d\xi \right) \partial_z \theta \right] \Theta\, dx dy dz + \int_\Omega F_\Theta \Theta dx dy dz.$$
$$(132)$$

We estimate now the terms on the right-hand side of (131) and (132). The first terms on the right-hand side of (131) can be estimated with the inequalities of Hölder, Ladyshenzkaya and Young

$$\left| \int_\Omega (\mathbf{V} \cdot \nabla)\mathbf{v} \cdot \mathbf{V}\, dx dy dz \right| \le c ||\nabla \mathbf{v}||_{\mathbf{L}^6} ||\mathbf{V}||_{\mathbf{L}^3} ||\mathbf{V}||_{\mathbf{L}^2}$$
$$\le c ||\nabla \mathbf{v}||_{\mathbf{H}^1} ||\mathbf{V}||_{\mathbf{H}^1} ||\mathbf{V}||_{\mathbf{L}^2} \le \frac{c}{2\epsilon_1} ||\nabla \mathbf{v}||_{\mathbf{H}^1}^2 ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{\epsilon_1}{2} ||\mathbf{V}||_{\mathbf{H}^1}^2 \qquad (133)$$
$$\le \left(\frac{c}{2\epsilon_1} ||\nabla \mathbf{v}||_{\mathbf{H}^1}^2 + \frac{\epsilon_1}{2} \right) ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{\epsilon_1}{2} (||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 + ||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2),$$

and for the first term on the right-hand side of (132) we have

$$\left| \int_\Omega (\mathbf{V} \cdot \nabla)\theta \cdot \Theta\, dx dy dz \right| \le c ||\nabla \theta||_{L^6} ||\Theta||_{L^3} ||\mathbf{V}||_{\mathbf{L}^2}$$
$$\le \frac{c}{2\epsilon_2} ||\nabla \theta||_{H^1}^2 ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{\epsilon_2}{2} ||\Theta||_{H^1}^2 \qquad (134)$$
$$\le \frac{c}{2\epsilon_2} ||\nabla \theta||_{H^1}^2 ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{\epsilon_2}{2} (||\Theta||_{L^2}^2 + ||\nabla \Theta||_{L^2}^2 + ||\partial_z \Theta||_{L^2}^2),$$

where the $\epsilon_i > 0$ are arbitrary and will be fixed later.

For the second term at the right-hand side of (131), we use the fact that the domain is cylindrical, $\Omega = M \times \{-h, 0\}$, and apply the triangle inequality and the inequalities of Cauchy-Schwarz and Hölder (with norms L^2, L^4, L^4) to obtain[3]

$$
\begin{aligned}
&\left| \int_\Omega \left[\left(\int_{-h}^z \nabla \cdot \mathbf{V}(x, y, \xi, t) \, d\xi \right) \partial_z \mathbf{v} \right] \cdot \mathbf{V} \, dx \, dy \, dz \right| \\
&\leq c \int_M \left(\int_{-h}^0 |\nabla \mathbf{V}(x, y, z, t)| \, dz \int_{-h}^0 |\partial_z \mathbf{v}(x, y, z, t)| \, |\mathbf{V}(x, y, z, t)| \, dz \right) dx \, dy \\
&\leq c \int_M \left(\left(\int_{-h}^0 |\nabla \mathbf{V}(x, y, z, t)| \, dz \right) \left(\int_{-h}^0 |\partial_z \mathbf{v}(x, y, z, t)|^2 \, dz \right)^{1/2} \right. \\
&\qquad \left. \times \left(\int_{-h}^0 |\mathbf{V}(x, y, z, t)|^2 \, dz \right)^{1/2} \right) dx \, dy \\
&\leq c \left(\int_M \left(\int_{-h}^0 |\nabla \mathbf{V}(x, y, z, t)| \, dz \right)^2 dx \, dy \right)^{1/2} \\
&\quad \left(\int_M \left(\int_{-h}^0 |\partial_z \mathbf{v}(x, y, z, t)|^2 \, dz \right)^2 dx \, dy \right)^{1/4} \left(\int_M \left(\int_{-h}^0 |\mathbf{V}(x, y, z, t)|^2 \, dz \right)^2 dx \, dy \right)^{1/4}.
\end{aligned}
\tag{135}
$$

For the first term on the right-hand side of (135), we obtain with Cauchy–Schwarz

$$
\left(\int_M \left(\int_{-h}^0 |\nabla \mathbf{V}(x, y, z, t)| \, dz \right)^2 dx \, dy \right)^{1/2} \leq c \|\nabla \mathbf{V}\|_{\mathbf{L}^2}.
\tag{136}
$$

For the second term, we obtain with the Minkowski inequality and the Ladyshenzkaya inequality in 2D

$$
\begin{aligned}
&\left(\int_M \left(\int_{-h}^0 |\partial_z \mathbf{v}(x, y, z, t)|^2 \, dz \right)^2 dx \, dy \right)^{1/2} \\
&\leq \int_{-h}^0 \left(\int_M |\partial_z \mathbf{v}(x, y, z, t)|^4 \, dx \, dy \right)^{1/2} dz \leq \|\partial_z \mathbf{v}\|_{\mathbf{L}^2} \|\nabla \partial_z \mathbf{v}\|_{\mathbf{L}^2}.
\end{aligned}
\tag{137}
$$

For the last term on the right-hand side of (135), it holds that

$$
\begin{aligned}
&\left(\int_M \left(\int_{-h}^0 |\mathbf{V}(x, y, z, t)|^2 \, dz \right)^2 dx \, dy \right)^{1/2} \leq \int_{-h}^0 \left(\int_M |\mathbf{V}(x, y, z, t)|^4 \, dx \, dy \right)^{1/2} dz \\
&\leq \int_{-h}^0 \|\mathbf{V}(\cdot, \cdot, z, t)\|_{\mathbf{L}^2} \|\nabla \mathbf{V}(\cdot, \cdot, z, t)\|_{\mathbf{L}^2} \, dz \leq c \|\mathbf{V}\|_{\mathbf{L}^2} \|\nabla \mathbf{V}\|_{\mathbf{L}^2}.
\end{aligned}
\tag{138}
$$

---

[3] The inequalities (135)–(139) apply a technique from (Cao and Titi 2007, see p. 264).

From (136)–(138), we obtain for (135)

$$
\left| \int_{\Omega} \left[\left(\int_{-h}^{z} \nabla \cdot \mathbf{V}(x, y, \xi, t) \, d\xi \right) \partial_z v \right] \cdot \mathbf{V} \, dxdydz \right|
$$
$$
\leq \frac{c}{\epsilon_3} ||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 ||\nabla \partial_z \mathbf{v}||_{\mathbf{L}^2}^2 ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{\epsilon_3}{2} ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2. \tag{139}
$$

The same argument yields the following estimate for the second term on the right-hand side of (132)

$$
\left| \int_{\Omega} \left[\int_{-h}^{z} \nabla \cdot \mathbf{V}(x, y, \xi, t) \, d\xi \, \partial_z \theta \right] \cdot \Theta \, dxdydz \right|
$$
$$
\leq \frac{c}{\epsilon_4} ||\partial_z \theta||_{L^2}^2 ||\nabla \partial_z \theta||_{L^2}^2 ||\Theta||_{L^2}^2 + \frac{\epsilon_4}{2} ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2. \tag{140}
$$

For the pressure gradient term in the velocity equation (129), the inequality of Cauchy–Schwarz and the Young inequality yield

$$
\left| \int_{\Omega} \nabla \left(\int_{-h}^{z} \Pi(x, y, \xi, t) d\xi \right) \mathbf{V} \, dxdydz \right| \leq h ||\Pi||_{L^2} \, ||\nabla \mathbf{V}||_{L^2}
$$
$$
\leq \frac{c}{2\epsilon_5} ||\Theta||_{L^2} + \frac{\epsilon_5}{2} ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2, \tag{141}
$$

where we have used the equation of state (32d) in the last step.

The forcing terms on the right-hand side of (129) and (130) are estimated as

$$
\int_{\Omega} \mathbf{V} \cdot F_{\mathbf{V}} \, dxdydz \leq ||\mathbf{V}||_{\mathbf{L}^2} ||F_{\mathbf{V}}||_{\mathbf{L}^2} \leq \frac{1}{2\epsilon_6} ||| F_{\mathbf{V}}||_{\mathbf{L}^2}^2 + \frac{\epsilon_6}{2} ||\mathbf{V}||_{\mathbf{L}^2}^2,
$$
$$
\text{and} \int_{\Omega} \Theta \, F_{\Theta} \, dxdydz \leq ||\Theta||_{L^2} ||F_{\Theta}||_{L^2} \leq \frac{1}{2\epsilon_7} ||F_{\Theta}||_{L^2}^2 + \frac{\epsilon_7}{2} ||\Theta||_{L^2}^2. \tag{142}
$$

Collecting the estimates (129)-(141) results in the following bounds for the linearized velocity equation

$$
\frac{1}{2} \frac{d}{dt} ||\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re_1} ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re_2} ||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2
$$
$$
\leq \left(\frac{c}{2\epsilon_1} ||\nabla \mathbf{v}||_{\mathbf{H}^1}^2 + \frac{c}{\epsilon_3} ||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 ||\nabla \partial_z \mathbf{v}||_{\mathbf{L}^2}^2 + \frac{\epsilon_1}{2} + \frac{\epsilon_6}{2} \right) ||\mathbf{V}||_{\mathbf{L}^2}^2 \tag{143}
$$
$$
+ \left(\frac{\epsilon_3}{2} + \frac{\epsilon_5}{2} \right) ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{c}{2\epsilon_5} ||\Theta||_{L^2} + \frac{1}{2\epsilon_6} ||F_{\mathbf{V}}||_{\mathbf{L}^2}^2,
$$

and for the temperature equation

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}||\Theta||^2_{L^2} + \frac{1}{Rt_1}||\nabla\Theta||^2_{L^2} + \frac{1}{Rt_2}||\partial_z\Theta||^2_{L^2} + k_\theta||\theta(z=0)||^2_{L^2} \\
&\leq \frac{c}{\epsilon_2}||\nabla\theta||^2_{H^1}||\mathbf{V}||^2_{\mathbf{L}^2} + \left(\frac{c}{\epsilon_4}||\partial_z\theta||^2_{L^2}||\nabla\partial_z\theta||^2_{L^2} + \frac{\epsilon_7}{2}\right)||\Theta||^2_{L^2} \qquad (144)\\
&+ \frac{\epsilon_2}{2}(||\Theta||^2_{L^2} + ||\nabla\Theta||^2_{L^2} + ||\partial_z\Theta||^2_{L^2}) + \frac{\epsilon_4}{2}||\nabla\mathbf{V}||^2_{\mathbf{L}^2} + \frac{1}{2\epsilon_7}||F_\Theta||^2_{L^2}.
\end{aligned}
$$

We combine now the two estimates (143) and (144). We define $\frac{1}{Re} := \min\{\frac{1}{Re_1}, \frac{1}{Re_2}\}$ and $\frac{1}{Rt} := \min\{\frac{1}{Rt_1}, \frac{1}{Rt_2}\}$ and choose the ϵ_i appropriately. This yields

$$
\begin{aligned}
&\frac{d}{dt}\left[||\mathbf{V}||^2_{\mathbf{L}^2} + ||\Theta||^2_{L^2}\right] + \frac{1}{Re}\left[||\nabla\mathbf{V}||^2_{\mathbf{L}^2} + ||\partial_z\mathbf{V}||^2_{\mathbf{L}^2}\right] \\
&\quad + \frac{1}{Rt}\left[||\nabla\Theta||^2_{L^2} + ||\partial_z\Theta||^2_{L^2}\right] + k_\theta||\theta(z=0)||^2_{L^2} \\
&\leq G\left(||\mathbf{V}||^2_{\mathbf{L}^2} + ||\Theta||^2_{L^2}\right) + c(||F_\mathbf{V}||^2_{\mathbf{L}^2} + ||F_\Theta||^2_{L^2}), \qquad (145)\\
&\text{where }\; G(s) := c(||\nabla\mathbf{v}(s)||^2_{\mathbf{H}^1} + ||\partial_z\mathbf{v}||^2_{\mathbf{L}^2}||\nabla\partial_z\mathbf{v}||^2_{\mathbf{L}^2} \\
&\qquad\qquad + ||\nabla\theta(s)||^2_{H^1} + ||\partial_z\theta||^2_{L^2}||\nabla\partial_z\theta||^2_{L^2} + 1).
\end{aligned}
$$

The Gronwall inequality implies

$$
\begin{aligned}
\left[||\mathbf{V}(t)||^2_{\mathbf{L}^2} + ||\Theta(t)||^2_{L^2}\right] &\leq \left[||\mathbf{V}_0||^2_{\mathbf{L}^2} + ||\Theta_0||^2_{L^2}\right]exp\left\{\int_0^t G(s)\,ds\right\} \\
&+ \int_0^t [||F_\mathbf{V}||^2_{\mathbf{L}^2} + ||F_\Theta||^2_{L^2}]exp\left\{\int_s^t G(y)\,dy\right\}ds.
\end{aligned}
\qquad (146)
$$

Since $(\mathbf{v}, \theta)$ is a strong solution this shows that $\mathbf{V}$ and Θ are bounded in $L^\infty([0, T], \mathbf{L}^2(\Omega))$.

Step 2: $L^\infty([0, T], \mathbf{H}^1(\Omega))$-bound on velocity

To prove the H^1-estimates for velocity we introduce the *hydrostatic Stokes operator* A (see Guillén-González et al. 2001; Kukavica and Ziane 2007). The hydrostatic Stokes operator is defined by $A\mathbf{v} := \mathbb{P}\triangle\mathbf{v}$ for $\mathbf{v} \in D(A)$, where $\mathbb{P}$ denotes the is the projection onto the space of square-integrable vector fields that are incompressible and satisfy the velocity boundary conditions (2)-(7). The domain of definition $D(A)$ of A is given by $D(A) := \{\mathbf{v} \in \mathbf{L}^2(\Omega) : \mathbf{v} \text{ and } A\mathbf{v} \text{ satisfy boundary conditions (2)-(7)},$ $\nabla_3 \cdot \mathbf{v} = 0, A\mathbf{v} \in \mathbf{L}^2(\Omega), \nabla_3 \cdot A\mathbf{v} = 0\}$. The Stokes operator satisfies $\langle A\mathbf{v}, \mathbf{v}\rangle_{\mathbf{L}^2} = \int_\Omega Re_1^{-1}\nabla\mathbf{v}\cdot\nabla\mathbf{v} + Re_2^{-1}\partial_z\mathbf{v}\cdot\partial_z\mathbf{v}\,dxdydz$.

For this a priori estimate, we treat the horizontal and vertical direction separately.

Step 2a: $L^\infty([0, T], L^2(\Omega))$-bound on ∇V

Taking the L^2-inner product of the linearized velocity equation (32a) with $A\mathbf{V}$ yields

$$
\begin{aligned}
\frac{1}{2}&\frac{\mathrm{d}}{\mathrm{d}t}||A^{1/2}\mathbf{V}||^2_{\mathbf{L}^2} + ||A\mathbf{V}||^2_{\mathbf{L}^2} \\
&= -\int_\Omega \left[((\mathbf{V}\cdot\nabla)\mathbf{v} + \mathbf{v}\cdot\nabla)\mathbf{V} \right. \\
&\quad\left. - \left(\nabla\cdot\int_{-h}^z \mathbf{v}(x,y,\xi,t)\,\mathrm{d}\xi\right)\partial_z\mathbf{V}\right]\cdot A\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad - \left(\nabla\cdot\int_{-h}^z \mathbf{V}(x,y,\xi,t)\,\mathrm{d}\xi\right)\partial_z\mathbf{v}\cdot AV\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega A\mathbf{V}\cdot F_\mathbf{V}\mathrm{d}x\mathrm{d}y\mathrm{d}z,
\end{aligned}
\tag{147}
$$

where we have used that the Stokes operator annihilates gradients and that the Coriolis term also vanishes. The forcing term is estimated by the inequalities of Cauchy–Schwarz and Young

$$
\int_\Omega A\mathbf{V}\cdot F_\mathbf{V}\mathrm{d}x\mathrm{d}y\mathrm{d}z \leq ||A\mathbf{V}||_{\mathbf{L}^2}||F_\mathbf{V}||_{\mathbf{L}^2} \leq \frac{\epsilon_0}{2}||A\mathbf{V}||^2_{\mathbf{L}^2} + \frac{1}{2\epsilon_0}||F_\mathbf{V}||^2_{\mathbf{L}^2}. \tag{148}
$$

For the first and second term on the right-hand side of (147) follows with the inequalities of Hölder, Ladyshenzkaya and Young

$$
\begin{aligned}
\left|\int_\Omega (\mathbf{V}\cdot\nabla)\mathbf{v}\cdot A\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\right| &\leq ||\mathbf{V}||_{\mathbf{L}^6}||\nabla\mathbf{v}||_{\mathbf{L}^3}||A\mathbf{V}||_{\mathbf{L}^2}, \\
&\leq c||\mathbf{V}||_{\mathbf{H}^1}||\nabla\mathbf{v}||_{\mathbf{H}^1}||A\mathbf{V}||_{\mathbf{L}^2}, \\
&\leq \frac{c}{2\epsilon_1}||\nabla\mathbf{v}||^2_{\mathbf{H}^1}||\mathbf{V}||^2_{\mathbf{H}^1} + \frac{\epsilon_1}{2}||A\mathbf{V}||^2_{\mathbf{L}^2},
\end{aligned}
\tag{149}
$$

and

$$
\begin{aligned}
\left|\int_\Omega (\mathbf{v}\cdot\nabla)\mathbf{V}\cdot A\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\right| &\leq ||\mathbf{v}||_{\mathbf{L}^6}||\nabla\mathbf{V}||_{\mathbf{L}^3}||A\mathbf{V}||_{\mathbf{L}^2}, \\
&\leq c||\mathbf{v}||_{\mathbf{H}^1}||\nabla\mathbf{V}||^{1/2}_{\mathbf{L}^2}||\Delta\mathbf{V}||^{1/2}_{\mathbf{L}^2}||A\mathbf{V}||_{\mathbf{L}^2}, \\
&\leq \frac{c}{2\epsilon_2}||\mathbf{v}||^2_{\mathbf{H}^1}||\nabla\mathbf{V}||_{\mathbf{L}^2}||A\mathbf{V}||_{\mathbf{L}^2} + \frac{\epsilon_2}{2}||A\mathbf{V}||^2_{\mathbf{L}^2}, \\
&\leq \frac{c}{4\epsilon_2\epsilon_3}||\mathbf{v}||^4_{\mathbf{H}^1}||\nabla\mathbf{V}||^2_{\mathbf{L}^2} + \left(\frac{\epsilon_2}{2} + \frac{\epsilon_3}{2}\right)||A\mathbf{V}||^2_{\mathbf{L}^2}.
\end{aligned}
\tag{150}
$$

For the third term on the right-hand side of (147) we obtain with Lemma 2.2 ($f = \Delta\mathbf{V}$, $g = \partial_z\mathbf{V}$) and the Young inequality

$$
\begin{aligned}
&\left| \int_\Omega \left(\nabla \cdot \int_{-h}^{z} \mathbf{v}(x, y, \xi, t)\, \mathrm{d}\xi \right) \partial_z \mathbf{V} A \mathbf{V}\, \mathrm{d}x\mathrm{d}y\mathrm{d}z \right| \\
&\leq C \|A\mathbf{V}\|_{\mathbf{L}^2} \|\mathbf{v}\|_{\mathbf{H}^1}^{1/2} \|\mathbf{v}\|_{\mathbf{H}^2}^{1/2} \|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^{1/2} \|\partial_z\mathbf{V}\|_{\mathbf{H}^1}^{1/2}. \\
&\leq \frac{\epsilon_4}{2} \|A\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{1}{2\epsilon_4} \|\mathbf{v}\|_{\mathbf{H}^1} \|\mathbf{v}\|_{\mathbf{H}^2} \|\partial_z\mathbf{V}\|_{\mathbf{L}^2} \|\partial_z\mathbf{V}\|_{\mathbf{H}^1} \\
&\leq \frac{\epsilon_4}{2} \|A\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{\epsilon_5}{2} \|\partial_z\mathbf{V}\|_{\mathbf{H}^1}^2 + \frac{1}{4\epsilon_4\epsilon_5} \|\mathbf{v}\|_{\mathbf{H}^1}^2 \|\mathbf{v}\|_{\mathbf{H}^2}^2 \|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2.
\end{aligned}
\tag{151}
$$

Applying Lemma 2.2 to the fourth term on the right-hand side of (147) ($f = \Delta V$, $g = \partial_z v$) yields together with the Young inequality

$$
\begin{aligned}
&\left| \left(\int_{-h}^{z} \nabla \cdot \mathbf{V}(x, y, \xi, t)\, \mathrm{d}\xi \right) \partial_z v \cdot A\mathbf{V}\, \mathrm{d}x\mathrm{d}y\mathrm{d}z \right| \\
&\leq C \|A\mathbf{V}\|_{\mathbf{L}^2} \|\mathbf{V}\|_{\mathbf{H}^1}^{1/2} \|\mathbf{V}\|_{\mathbf{H}^2}^{1/2} \|\partial_z v\|_{\mathbf{L}^2}^{1/2} \|\partial_z v\|_{\mathbf{H}^1}^{1/2} \\
&\leq \frac{\epsilon_6}{2} \|A\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{c}{2\epsilon_6} \|\mathbf{V}\|_{\mathbf{H}^1} \|V\|_{\mathbf{H}^2} \|\partial_z v\|_{\mathbf{L}^2} \|\partial_z v\|_{\mathbf{H}^1} \\
&\leq \frac{\epsilon_6}{2} \|A\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{\epsilon_7}{2} \|\mathbf{V}\|_{\mathbf{H}^2}^2 + \frac{c}{2\epsilon_6^2\epsilon_7} \|\mathbf{V}\|_{\mathbf{H}^1}^2 \|\partial_z v\|_{\mathbf{L}^2}^2 \|\partial_z v\|_{\mathbf{H}^1}^2.
\end{aligned}
\tag{152}
$$

Summarizing our estimates yields

$$
\begin{aligned}
&\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \|\nabla\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{1}{Re} \|\Delta\mathbf{V}\|_{\mathbf{L}^2}^2 + \left(\frac{1}{Re_1} + \frac{1}{Re_2} \right) \|\nabla\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2 \\
&\leq \left(\frac{c}{2\epsilon_1} \|\nabla\mathbf{v}\|_{\mathbf{H}^1}^2 + \frac{c}{4\epsilon_6\epsilon_7} \|\partial_z v\|_{\mathbf{L}^2}^2 \|\partial_z v\|_{\mathbf{H}^1}^2 \right) \|\mathbf{V}\|_{\mathbf{H}^1}^2 + \frac{c}{4\epsilon_2\epsilon_3} \|\mathbf{v}\|_{\mathbf{H}^1}^4 \|\nabla\mathbf{V}\|_{\mathbf{L}^2}^2 \\
&\quad + \left(\frac{\epsilon_0}{2} + \frac{\epsilon_1}{2} + \frac{\epsilon_2}{2} + \frac{\epsilon_3}{2} + \frac{\epsilon_4}{2} + \frac{\epsilon_6}{2} \right) \|\Delta\mathbf{V}\|_{\mathbf{L}^2}^2 \\
&\quad + \frac{\epsilon_5}{2} \left(\|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2 + \|\nabla\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2 + \|\partial_{zz}^2\mathbf{V}\|_{\mathbf{L}^2}^2 \right) \\
&\quad + \frac{c}{2\epsilon_4\epsilon_5} \|\mathbf{v}\|_{\mathbf{H}^1}^2 \|\mathbf{v}\|_{\mathbf{H}^2}^2 \|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{\epsilon_7}{2} \|\mathbf{V}\|_{\mathbf{H}^2}^2 + \frac{1}{2\epsilon_0} \|F_\mathbf{V}\|_{\mathbf{L}^2}^2.
\end{aligned}
\tag{153}
$$

We need a bound on $\|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^2$ in (153). This is done in the next step before, we complete the $L^\infty([0, T], \mathbf{H}^1(\Omega))$ bound on the velocity.

Step 2b: $L^\infty([0, T], L^2(\Omega))$-bound on $\partial_z V$.

The vertical derivative $\mathbf{U} := \partial_z \mathbf{V}$ of the linearized velocity satisfies the following equation

$$
\begin{aligned}
\frac{\partial \mathbf{U}}{\partial t} &+ (\mathbf{U} \cdot \nabla)\mathbf{v} + (V \cdot \nabla)\partial_z \mathbf{v} + (\partial_z \mathbf{v} \cdot \nabla)\mathbf{V} + (\mathbf{v} \cdot \nabla)\mathbf{U} \\
&- \nabla \cdot \mathbf{V}\partial_z \mathbf{v} - \left(\nabla \cdot \int_{-h}^{z} \mathbf{V}(x, y, \xi, t)\,\mathrm{d}\xi\right)\partial_{zz}^2 \mathbf{v} \\
&- (\nabla \cdot \mathbf{v})\mathbf{U} - \left(\nabla \cdot \int_{-h}^{z} \mathbf{v}(x, y, \xi, t)\,\mathrm{d}\xi\right)\partial_z \mathbf{U} \\
&- \frac{1}{Re_1}\triangle \mathbf{U} - \frac{1}{Re_2}\partial_{zz}^2 \mathbf{U} + g\nabla\partial_z \Pi + f\vec{k} \times \mathbf{U} = \partial_z F_{\mathbf{V}}.
\end{aligned}
\tag{154}
$$

Taking the L^2-scalar product of (154) with $\mathbf{U}$ yields

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}&||\mathbf{U}||_{\mathbf{L}^2}^2 + \frac{1}{Re_1}||\nabla \mathbf{U}||_{\mathbf{L}^2}^2 + \frac{1}{Re_2}||\partial_z \mathbf{U}||_{\mathbf{L}^2}^2 \\
&= -\int_\Omega [(\mathbf{U} \cdot \nabla)\mathbf{v}] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z - \int_\Omega \left[(V \cdot \nabla)\partial_z \mathbf{v}\right] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad - \int_\Omega \left[(\partial_z \mathbf{v} \cdot \nabla)\mathbf{V}\right] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z - \int_\Omega [(\mathbf{v} \cdot \nabla)\mathbf{U}] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad + \int_\Omega [(\nabla \cdot \mathbf{v})\mathbf{U}] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + \int_\Omega \left[\nabla \cdot \int_{-h}^{z} \mathbf{v}(x, y, \xi, t)\,\mathrm{d}\xi\partial_z \mathbf{U}\right] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad + \int_\Omega \left[\nabla \cdot \int_{-h}^{z} \mathbf{V}(x, y, \xi, t)\,\mathrm{d}\xi\partial_{zz}^2 \mathbf{v}\right] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z - \int_\Omega \left[g\nabla\partial_z \Pi\right] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&\quad + \int_\Omega \left[\partial_z F_V\right] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z.
\end{aligned}
\tag{155}
$$

For the first term on the right-hand side of (155) we obtain with the inequalities of Hölder, Ladyshenzkaya and Young

$$
\begin{aligned}
\left|\int_\Omega [(\mathbf{U} \cdot \nabla)\mathbf{v}] \cdot \mathbf{U}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\right| &\leq ||\nabla\mathbf{v}||_{\mathbf{L}^2}||\mathbf{U}||_{\mathbf{L}^6}||\mathbf{U}||_{\mathbf{L}^3} \\
&\leq ||\nabla\mathbf{v}||_{\mathbf{L}^2}||\mathbf{U}||_{\mathbf{L}^6}||\mathbf{U}||_{\mathbf{L}^2}^{1/2}||\nabla\mathbf{U}||_{\mathbf{L}^2}^{1/2} \\
&\leq \frac{C_0}{2\epsilon_8}||\nabla\mathbf{v}||_{\mathbf{L}^6}^2||\mathbf{U}||_{\mathbf{L}^2}||\nabla\mathbf{U}||_{\mathbf{L}^2} + \frac{\epsilon_8}{2}||\mathbf{U}||_{\mathbf{H}^1}^2 \\
&\leq \frac{C_0}{4\epsilon_8\epsilon_9}||\nabla\mathbf{v}||_{\mathbf{L}^6}^4||\mathbf{U}||_{\mathbf{L}^2}^2 + \frac{\epsilon_9}{2}||\nabla\mathbf{U}||_{\mathbf{L}^2}^2 + \frac{\epsilon_8}{2}(||\mathbf{U}||_{\mathbf{L}^2}^2 + ||\nabla\mathbf{U}||_{\mathbf{L}^2}^2 + ||\partial_z\mathbf{U}||_{\mathbf{L}^2}^2).
\end{aligned}
\tag{156}
$$

For the third term on the right-hand side of (155) follows the Hölder and Young inequalities

$$
\begin{aligned}
\left| \int_\Omega \left[(\partial_z \mathbf{v} \cdot \nabla) \mathbf{V} \right] \cdot \mathbf{U} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z \right| &\leq ||\partial_z \mathbf{v}||_{\mathbf{L}^6} ||\nabla \mathbf{V}||_{\mathbf{L}^2} ||\mathbf{U}||_{\mathbf{L}^3} \\
&\leq \frac{c}{2\epsilon_{10}} ||\partial_z \mathbf{v}||^2_{\mathbf{H}^1} ||\mathbf{U}||_{\mathbf{L}^2} ||\nabla \mathbf{U}||_{\mathbf{L}^2} + \frac{\epsilon_{10}}{2} ||\nabla \mathbf{V}||^2_{\mathbf{L}^2} \\
&\leq \frac{c}{4\epsilon_{10}\epsilon_{11}} ||\partial_z v||^4_{\mathbf{H}^1} ||\mathbf{U}||^2_{\mathbf{L}^2} + \frac{\epsilon_{11}}{2} ||\nabla \mathbf{U}||_{\mathbf{L}^2} + \frac{\epsilon_{10}}{2} ||\nabla \mathbf{V}||^2_{\mathbf{L}^2}.
\end{aligned}
\tag{157}
$$

Analogously follows for the fifth term

$$
\begin{aligned}
\left| \int_\Omega \left[(\nabla \cdot \mathbf{v}) \mathbf{U} \right] \cdot \mathbf{U} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z \right| &\leq ||\nabla \mathbf{v}||_{\mathbf{L}^6} ||\mathbf{U}||_{\mathbf{L}^2} ||\mathbf{U}||_{\mathbf{L}^3} \\
&\leq C_0 ||\nabla \mathbf{v}||_{\mathbf{H}^1} ||\mathbf{U}||^{3/2}_{\mathbf{L}^2} ||\nabla \mathbf{U}||^{1/2}_{\mathbf{L}^2} \leq \frac{C_0}{2\epsilon_{12}} ||\nabla \mathbf{v}||^4_{\mathbf{H}^1} ||\mathbf{U}||^2_{\mathbf{L}^2} + \frac{\epsilon_{12}}{2} ||\nabla \mathbf{U}||^2_{\mathbf{L}^2}.
\end{aligned}
\tag{158}
$$

We observe that

$$
\int_\Omega \left[(\mathbf{V} \cdot \nabla)\partial_z \mathbf{v} - \left(\nabla \cdot \int_{-h}^z \mathbf{V}(x,y,\xi,t) \, \mathrm{d}\xi \right) \partial^2_{zz} \mathbf{v} \right] \cdot \mathbf{U} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z = 0,
$$

and

$$
\int_\Omega \left[(\mathbf{v} \cdot \nabla)\mathbf{U} - \left(\nabla \cdot \int_{-h}^z \mathbf{v}(x,y,\xi,t) \, \mathrm{d}\xi \right) \partial_z \mathbf{U} \right] \cdot \mathbf{U} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z = 0,
\tag{159}
$$

and on the right-hand side of (155) the second and seventh as well as the fourth and sixth term cancel. The pressure gradient (the eight term in (155)) is bounded as follows

$$
\int_\Omega \left[g \nabla \partial_z \Pi \right] \cdot \mathbf{U} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z \leq ||\partial_z \Pi||_{L^2} ||\nabla \mathbf{U}||_{\mathbf{L}^2} \leq \frac{1}{2\epsilon_{13}} ||\partial_z \Theta||^2_{L^2} + \frac{\epsilon_{13}}{2} ||\nabla \mathbf{U}||^2_{\mathbf{L}^2},
\tag{160}
$$

where we have used (32d). For the forcing term, we obtain

$$
\int_\Omega \partial_z F_\mathbf{V} \cdot \mathbf{U} \, \mathrm{d}x\mathrm{d}y\mathrm{d}z \leq ||F_\mathbf{V}||_{\mathbf{L}^2} ||\partial_z \mathbf{U}||_{\mathbf{L}^2} \leq \frac{1}{2\epsilon_{14}} ||F_\mathbf{V}||^2_{\mathbf{L}^2} + \frac{\epsilon_{14}}{2} ||\partial_z \mathbf{U}||^2_{\mathbf{L}^2}.
\tag{161}
$$

In summary, we obtain from (155)–(161) the following estimate

$$
\begin{aligned}
&\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} ||\partial_z \mathbf{V}||^2_{\mathbf{L}^2} + \frac{1}{Re_1} ||\nabla \partial_z \mathbf{V}||^2_{\mathbf{L}^2} + \frac{1}{Re_2} ||\partial^2_{zz} \mathbf{V}||^2_{\mathbf{L}^2} \\
&\leq c \left(\frac{1}{4\epsilon_{10}\epsilon_{11}} ||\partial_z \mathbf{v}||^4_{\mathbf{H}^1} + \left(\frac{1}{4\epsilon_8\epsilon_9} + \frac{1}{2\epsilon_{12}} \right) ||\nabla \mathbf{v}||^4_{\mathbf{H}^1} \right) ||\partial_z \mathbf{V}||^2_{\mathbf{L}^2} \\
&\quad + \left(\frac{\epsilon_9}{2} + \frac{\epsilon_{11}}{2} + \frac{\epsilon_{12}}{2} + \frac{\epsilon_{13}}{2} \right) ||\nabla \partial_z \mathbf{V}||^2_{\mathbf{L}^2}
\end{aligned}
$$

$$+ \frac{\epsilon_8}{2} C_0 (||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2)$$

$$+ \frac{\epsilon_{10}}{2} ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{\epsilon_{14}}{2} ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{2\epsilon_{13}} ||\partial_z \Theta||_{L^2}^2 + \frac{1}{2\epsilon_{14}} ||F_{\mathbf{V}}||_{\mathbf{L}^2}^2. \quad (162)$$

Step 2c: Completing the $L^\infty([0, T], \mathbf{H}^1(\Omega))$-bound on velocity.

For the $L^2([0, T], \mathbf{H}^1(\Omega))$-estimate of velocity follows from (153) and (162)

$$\begin{aligned}
\frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} (||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 &+ ||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2) + \frac{1}{Re} ||\Delta \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re} ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re} ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 \\
&\leq c \left(\frac{1}{2\epsilon_1} ||\nabla \mathbf{v}||_{\mathbf{H}^1}^2 + \frac{1}{4\epsilon_6 \epsilon_7} ||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 ||\partial_z \mathbf{v}||_{\mathbf{H}^1}^2 \right) ||\mathbf{V}||_{\mathbf{H}^1}^2 \\
&+ c \left(\frac{1}{4\epsilon_2 \epsilon_3} ||\mathbf{v}||_{\mathbf{H}^1}^4 + \frac{\epsilon_{10}}{2} \right) ||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 \\
&+ c \left(\frac{1}{4\epsilon_{10}\epsilon_{11}} ||\partial_z \mathbf{v}||_{\mathbf{H}^1}^4 + (\frac{1}{4\epsilon_4 \epsilon_5} ||\mathbf{v}||_{\mathbf{H}^1}^2 ||\mathbf{v}||_{\mathbf{H}^2}^2 + \frac{1}{4\epsilon_8 \epsilon_9} + \frac{1}{2\epsilon_{12}}) ||\nabla \mathbf{v}||_{\mathbf{H}^1}^4 \right) ||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2 \\
&+ \left(\frac{\epsilon_0}{2} + \frac{\epsilon_1}{2} + \frac{\epsilon_2}{2} + \frac{\epsilon_3}{2} + \frac{\epsilon_4}{2} + \frac{\epsilon_6}{2} \right) ||\Delta \mathbf{V}||_{\mathbf{L}^2}^2 \\
&+ \left(\frac{\epsilon_5}{2} + \frac{\epsilon_8}{2} \right) \left(||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 \right) + \frac{\epsilon_{14}}{2} ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 \\
&+ \frac{\epsilon_7}{2} \left(||\mathbf{V}||_{\mathbf{L}^2}^2 + ||\mathbf{V}||_{\mathbf{H}^1}^2 + ||\Delta \mathbf{V}||_{\mathbf{L}^2}^2 + ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 + ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 \right) \\
&+ \left(\frac{\epsilon_9}{2} + \frac{\epsilon_{11}}{2} + \frac{\epsilon_{12}}{2} + \frac{\epsilon_{13}}{2} \right) ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 \\
&+ \frac{1}{2\epsilon_{13}} ||\partial_z \Theta||_{L^2}^2 + \quad (\frac{1}{2\epsilon_0} + \frac{1}{2\epsilon_{14}}) ||F_{\mathbf{V}}||_{\mathbf{L}^2}^2.
\end{aligned} \quad (163)$$

We choose the parameters ϵ_i such that the terms with second-order derivatives on the right-hand side of (163) can be compensated with the corresponding terms on the left-hand side. This yields

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} (||\nabla \mathbf{V}||_{\mathbf{L}^2}^2 &+ ||\partial_z \mathbf{V}||_{\mathbf{L}^2}^2) + \frac{1}{Re} ||\Delta \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re} ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re} ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 \\
&\leq c \left(||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 + ||\partial_z \mathbf{v}||_{\mathbf{H}^1}^4 + ||\mathbf{v}||_{\mathbf{H}^1}^4 + ||\partial_z \mathbf{v}||_{\mathbf{H}^1}^4 + ||\mathbf{v}||_{\mathbf{H}^1}^2 ||\mathbf{v}||_{\mathbf{H}^2}^2 + ||\nabla \mathbf{v}||_{\mathbf{H}^1}^4 + 1 \right) \times \\
&\times ||\mathbf{V}||_{\mathbf{H}^1}^2 + c ||\Theta||_{\mathbf{H}^1}^2 + c ||F_{\mathbf{V}}||_{\mathbf{L}^2}^2.
\end{aligned} \quad (164)$$

Combining this with the $L^\infty([0, T], L^2)$-estimate (143) for velocity, we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} ||\mathbf{V}||_{\mathbf{H}^1}^2 &+ \frac{1}{Re} ||\Delta \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re} ||\nabla \partial_z \mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re} ||\partial_{zz}^2 \mathbf{V}||_{\mathbf{L}^2}^2 \\
&\leq c K_1 \left(||V||_{\mathbf{H}^1}^2 + ||\Theta||_{\mathbf{H}^1}^2 + ||F_{\mathbf{V}}||_{\mathbf{L}^2}^2 \right),
\end{aligned} \quad (165)$$

where

$$\begin{aligned}
K_1(t) :=& ||\partial_z \mathbf{v}||_{\mathbf{L}^2}^2 + ||\partial_z \mathbf{v}||_{\mathbf{H}^1}^2 + ||\mathbf{v}||_{\mathbf{H}^1}^4 + ||\partial_z \mathbf{v}||_{\mathbf{H}^1}^4 \\
&+ ||\mathbf{v}||_{\mathbf{H}^1}^2 ||\mathbf{v}||_{\mathbf{H}^2}^2 + ||\nabla \mathbf{v}||_{\mathbf{H}^1}^4 + 1.
\end{aligned}$$

Step 3: $L^\infty([0, T], H^1(\Omega))$-bound on temperature.

After taking the inner product of the temperature equation (32c) with $-\Delta_3\Theta :=$ $-\Delta\Theta - \partial_{zz}^2\Theta$,

$$
\begin{aligned}
&\frac{1}{2}\frac{d}{dt}[||\nabla\Theta||_{L^2}^2 + ||\partial_z\Theta||_{L^2}^2 + k_\theta||\nabla\Theta(z=0)||_{L^2}^2] + \frac{1}{Rt_1}||\Delta\Theta||_{L^2}^2 + \frac{1}{Rt_2}||\partial_{zz}^2\Theta||_{L^2}^2 \\
&\quad + \left(\frac{1}{Rt_1} + \frac{1}{Rt_2}\right)(||\nabla\partial_z\Theta||_{L^2}^2 + k_\theta||\nabla\Theta(z=0)||_{L^2}^2) \\
&= -\int_\Omega [(\mathbf{v}\cdot\nabla)\Theta + (\mathbf{V}\cdot\nabla)\theta \\
&\qquad - \left(\int_{-h}^z \nabla\cdot\mathbf{v}(x,y,\xi,t)\,d\xi\right)\partial_z\Theta\right](\Delta\Theta + \partial_{zz}^2\Theta)\,dxdydz \\
&\qquad - \int_\Omega\left(\int_{-h}^z \nabla\cdot\mathbf{V}(x,y,\xi,t)\,d\xi\right)\partial_z\theta\,(\Delta\Theta + \partial_{zz}^2\Theta)\,dxdydz \\
&\qquad - \int_\Omega(\Delta\Theta + \partial_{zz}^2\Theta)F_\Theta\,dxdydz.
\end{aligned}
\tag{166}
$$

We estimate the first two terms on the right-hand side of (166) with the inequalities of Hölder, Ladyshenzkaya and Young

$$
\begin{aligned}
&|\int_\Omega (\mathbf{v}\cdot\nabla)\Theta(\Delta\Theta + \partial_{zz}^2\Theta)\,dxdydz| \\
&\quad\le \left(\frac{c}{4\epsilon_0\epsilon_2} + \frac{c}{4\epsilon_1^2\epsilon_3}\right)||\mathbf{v}||_{\mathbf{H}^1}^4||\nabla\Theta||_{L^2}^2 + \left(\frac{\epsilon_2}{2} + \frac{\epsilon_3}{2}\right)||\nabla\Theta||_{H^1}^2 \\
&\qquad + \frac{\epsilon_0}{2}||\Delta\Theta||_{L^2}^2 + \frac{\epsilon_1}{2}||\partial_{zz}^2\Theta||_{L^2}^2,
\end{aligned}
\tag{167}
$$

and

$$
\begin{aligned}
&|\int_\Omega (\mathbf{V}\cdot\nabla)\theta(\Delta\Theta + \partial_{zz}^2\Theta)\,dxdydz| \\
&\quad\le ||\nabla\theta||_{L^6}||\mathbf{V}||_{\mathbf{L}^3}||\Delta\Theta||_{L^2} + ||\nabla\theta||_{L^6}||\mathbf{V}||_{\mathbf{L}^3}||\partial_{zz}^2\Theta||_{L^2} \\
&\quad\le \left(\frac{c}{4\epsilon_5\epsilon_7} + \frac{c}{2\epsilon_6\epsilon_8}\right)||\theta||_{H^1}^4||\mathbf{V}||_{\mathbf{L}^2}^2 + \left(\frac{\epsilon_7}{2} + \frac{\epsilon_8}{2}\right)||\mathbf{V}||_{\mathbf{H}^1}^2 \\
&\qquad + \frac{\epsilon_4}{2}||\Delta\Theta||_{L^2}^2 + \frac{\epsilon_5}{2}||\partial_{zz}^2\Theta||_{L^2}^2.
\end{aligned}
\tag{168}
$$

For the third term in (166), we find analogously to (151)

$$
\begin{aligned}
&\left|\int_\Omega\left(\int_{-h}^z \nabla\cdot\mathbf{v}(x,y,\xi,t)\,d\xi\right)\partial_z\Theta(\Delta\Theta + \partial_{zz}^2\Theta)\,dxdydz\right| \\
&\quad\le \frac{\epsilon_9}{2}||\Delta\Theta||_{L^2}^2 + \frac{\epsilon_{11}}{2}||\partial_{zz}^2\Theta||_{L^2}^2 + \left(\frac{\epsilon_{12}}{2} + \frac{\epsilon_{10}}{2}\right)||\partial_z\Theta||_{H^1}^2
\end{aligned}
$$

$$+ \left(\frac{1}{2\epsilon_9\epsilon_{10}} + \frac{1}{2\epsilon_{11}\epsilon_{12}} \right) ||\mathbf{v}||^2_{\mathbf{H}^1} ||\mathbf{v}||^2_{\mathbf{H}^2} ||\partial_z\Theta||^2_{L^2}. \tag{169}$$

and for the fourth term

$$\left| \int_\Omega \left(\int_{-h}^z \nabla \cdot \mathbf{V}(x, y, \xi, t)\,d\xi \right) \partial_z\theta(\Delta\Theta + \partial^2_{zz}\Theta)\,dxdydz \right|$$
$$\leq \frac{\epsilon_{13}}{2}||\Delta\Theta||^2_{L^2} + \frac{\epsilon_{15}}{2}||\partial^2_{zz}\Theta||^2_{L^2} + \left(\frac{\epsilon_{16}}{2} + \frac{\epsilon_{14}}{2} \right) ||\Delta\Theta||^2_{L^2} \tag{170}$$
$$+ \left(\frac{1}{2\epsilon_{13}\epsilon_{14}} + \frac{1}{2\epsilon_{15}\epsilon_{16}} \right) ||\mathbf{V}||^2_{\mathbf{H}^1} ||\mathbf{V}||^2_{\mathbf{H}^2} ||\partial_z\theta||^2_{L^2}.$$

For the forcing, we obtain with the inequalities of Cauchy–Schwarz and Young

$$\left| \int_\Omega F_\Theta(\Delta\Theta + \partial^2_{zz}\Theta)\,dxdydz \right| \leq ||F_\Theta||_{L^2}||\Delta\Theta||_{L^2} + ||F_\Theta||_{L^2}||\partial^2_{zz}\Theta||_{L^2}$$
$$\leq \frac{\epsilon_{17}}{2}||\Delta\Theta||^2_{L^2} + \frac{\epsilon_{18}}{2}|||\partial^2_{zz}\Theta||^2_{L^2} + \left(\frac{1}{2\epsilon_{17}} + \frac{1}{2\epsilon_{18}} \right) ||F_\Theta||^2_{L^2}. \tag{171}$$

Collecting the estimates (167)-(171), we obtain for (166)

$$\frac{1}{2}\frac{d}{dt}[||\nabla\Theta||^2_{L^2} + ||\partial_z\Theta||^2_{L^2} + k_\theta||\nabla\Theta(z=0)||^2_{L^2}] + \frac{1}{Rt}||\Delta\Theta||^2_{L^2} + \frac{1}{Rt}||\partial^2_{zz}\Theta||^2_{L^2}$$
$$+ \frac{1}{Rt}||\nabla\partial_z\Theta||^2_{L^2} + k_\theta||\nabla\Theta(z=0)||^2_{L^2}$$
$$\leq \left(\frac{\epsilon_0}{2} + \frac{\epsilon_4}{2} + \frac{\epsilon_9}{2} + \frac{\epsilon_{13}}{2} + \frac{\epsilon_{14}}{2} + \frac{\epsilon_{16}}{2} + \frac{\epsilon_{17}}{2} \right) ||\Delta\Theta||^2_{L^2}$$
$$+ \left(\frac{\epsilon_1}{2} + \frac{\epsilon_5}{2} + \frac{\epsilon_{11}}{2} + \frac{\epsilon_{15}}{2} + \frac{\epsilon_{18}}{2} \right) ||\partial^2_{zz}\Theta||^2_{L^2}$$
$$+ \left(\frac{c}{4\epsilon_0\epsilon_2} + \frac{c}{4\epsilon_1^2\epsilon_3} \right) ||\mathbf{v}||^4_{\mathbf{H}^1} ||\nabla\Theta||^2_{L^2} + \left(\frac{\epsilon_2}{2} + \frac{\epsilon_3}{2} \right) \left(||\Delta\Theta||^2_{L^2} + ||\nabla\partial_z\Theta||^2_{L^2} \right)$$
$$+ \left(\frac{c}{4\epsilon_5\epsilon_7} + \frac{c}{2\epsilon_6\epsilon_8} \right) ||\theta||^4_{H^1} ||\mathbf{V}||^2_{L^2} + \left(\frac{\epsilon_7}{2} + \frac{\epsilon_8}{2} \right) ||\mathbf{V}||^2_{\mathbf{H}^1}$$
$$+ \left(\frac{\epsilon_{12}}{2} + \frac{\epsilon_{10}}{2} \right) \left(||\nabla\partial_z\Theta||^2_{L^2} + ||\partial^2_{zz}\Theta||^2_{L^2} \right)$$
$$+ \left(\frac{1}{2\epsilon_9\epsilon_{10}} + \frac{1}{2\epsilon_{11}\epsilon_{12}} \right) ||\mathbf{v}||^2_{\mathbf{H}^1} ||\mathbf{v}||^2_{H^2} ||\partial_z\Theta||^2_{L^2}$$
$$+ \left(\frac{1}{2\epsilon_{13}\epsilon_{14}} + \frac{1}{2\epsilon_{15}\epsilon_{16}} \right) ||\mathbf{V}||^2_{\mathbf{H}^1} ||\mathbf{V}||^2_{\mathbf{H}^2} ||\partial_z\theta||^2_{L^2} + (\frac{1}{2\epsilon_{17}} + \frac{1}{2\epsilon_{18}})||F_\Theta||^2_{L^2}$$
$$\tag{172}$$

Specifying the free parameters ϵ_i appropriately yields

$$
\frac{d}{dt}[||\nabla\Theta||_{L^2}^2 + ||\partial_z\Theta||_{L^2}^2 + k_\theta||\nabla\Theta(z=0)||_{L^2}^2] + \frac{1}{Rt}||\Delta\Theta||_{L^2}^2 + \frac{1}{Rt}||\partial_{zz}^2\Theta||_{L^2}^2
$$
$$
+ \frac{1}{Rt}(||\nabla\partial_z\Theta||_{L^2}^2 + k_\theta||\nabla\Theta(z=0)||_{L^2}^2) \tag{173}
$$
$$
\leq K_2||\mathbf{V}||_{\mathbf{H}^1}^2 + K_1||\Theta||_{H^1}^2 + c||F_\Theta||_{L^2}^2,
$$

where $K_2 := ||\mathbf{v}||_{\mathbf{H}^1}^4 + ||\mathbf{v}||_{\mathbf{H}^1}^2 + ||\mathbf{v}||_{\mathbf{H}^2}^2 + ||\theta||_{H^1}^4 + 1$,

and where K_1 was defined in (165). We combine the velocity estimate (165) and the temperature estimates (144), (173) and obtain

$$
\frac{d}{dt}[||\mathbf{V}||_{\mathbf{H}^1}^2 + ||\Theta||_{H^1}^2] + \frac{1}{Re}||\Delta\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re}||\nabla\partial_z\mathbf{V}||_{\mathbf{L}^2}^2 + \frac{1}{Re}||\partial_{zz}^2\mathbf{V}||_{\mathbf{L}^2}^2
$$
$$
+ \frac{1}{Rt}||\Delta\Theta||_{L^2}^2 + \frac{1}{Rt}||\partial_{zz}^2\Theta||_{L^2}^2 + \frac{1}{Rt}||\nabla\partial_z\Theta||_{L^2}^2 \tag{174}
$$
$$
\leq (K_1 + K_2)(||\mathbf{V}||_{\mathbf{H}^1}^2 + ||\Theta||_{H^1}^2) + c(||F_{\mathbf{V}}||_{\mathbf{L}^2}^2 + ||F_\Theta||_{L^2}^2).
$$

Gronwall's inequality implies

$$
||\mathbf{V}(t)||_{\mathbf{H}^1}^2 + ||\Theta(t)||_{H^1}^2 \leq [||\mathbf{V}_0||_{\mathbf{H}^1}^2 + ||\Theta_0||_{H^1}^2]\exp\{\int_0^t K_1(s) + K_2(s)\,ds\}
$$
$$
+ c\int_0^t ||F_{\mathbf{V}}(s)||_{\mathbf{L}^2}^2 + ||F_\Theta(s)||_{L^2}^2)\exp\{\int_0^s K_1(y) + K_2(y)\,dy\}\,ds \tag{175}
$$

where K_1, K_2 are bounded functions of s. This shows that $(\mathbf{V}, \Theta)$ are bounded in $L^\infty([0, T], \mathbf{H}^1)$.

After integrating (174) over the time interval $(0, t)$, $t \in (0, T]$ we get

$$
\int_0^t \frac{1}{Re}||\Delta\mathbf{V}(s)||_{\mathbf{L}^2}^2 + \frac{1}{Re}||\nabla\partial_z\mathbf{V}(s)||_{\mathbf{L}^2}^2 + \frac{1}{Re}||\partial_{zz}^2\mathbf{V}(s)||_{\mathbf{L}^2}^2
$$
$$
+ \frac{1}{Rt}||\Delta\Theta(s)||_{L^2}^2 + \frac{1}{Rt}||\partial_{zz}^2\Theta(s)||_{L^2}^2 + \frac{1}{Rt}||\nabla\partial_z\Theta(s)||_{L^2}^2 ds \tag{176}
$$
$$
\leq [||\mathbf{V}_0||_{\mathbf{H}^1}^2 + ||\Theta_0||_{H^1}^2] + \int_0^t (K_1(s) + K_2(s))(||\mathbf{V}(s)||_{\mathbf{H}^1}^2 + ||\Theta(s)||_{H^1}^2)
$$
$$
+ c(||F_{\mathbf{V}}(s)||_{\mathbf{L}^2}^2 + ||F_\Theta(s)||_{L^2}^2)\,ds.
$$

This proves that $(\mathbf{V}, \Theta)$ are bounded in $L^2([0, T], \mathbf{H}^2(\Omega))$.

Step 4: Time Derivative.

We take the L^2-scalar product of (32) with $\partial_t \Psi$. For the linearized velocity equation, we obtain

$$
\begin{aligned}
\left\|\frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t}\right\|_{\mathbf{L}^2}^2 &\leq \frac{1}{Re_1}\Big|\int_\Omega \Delta\mathbf{V}\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\Big| + \frac{1}{Re_2}\Big|\int_\Omega \partial_{zz}^2\mathbf{V}\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\Big| \\
&+ \Big|\int_\Omega \Big[(\mathbf{v}\cdot\nabla)\mathbf{V} + (\mathbf{V}\cdot\nabla)\mathbf{v} - \Big(\int_{-h}^z \nabla\cdot\mathbf{v}(x,y,\xi,t)\,\mathrm{d}\xi\Big)\partial_z\mathbf{V}\Big]\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\Big| \\
&+ \Big|\Big(\int_{-h}^z \nabla\cdot\mathbf{V}(x,y,\xi,t)\,\mathrm{d}\xi\Big)\partial_z\mathbf{v}\cdot\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&+ \Big|\int_\Omega \nabla\Big(\int_{-h}^z \Pi(x,y,\xi,t)\mathrm{d}\xi\Big)\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z + \Big|\int_\Omega \partial_t\mathbf{V}\cdot F_V\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\Big|,
\end{aligned}
\tag{177}
$$

where the terms involving the Coriolis force and the surface pressure vanish. We note that the sum of the third and fifth term on the right-hand side of (177) vanishes

$$
\int_\Omega \Big[(\mathbf{v}\cdot\nabla)\mathbf{V} - \Big(\int_{-h}^z \nabla\cdot\mathbf{v}(x,y,\xi,t)\,\mathrm{d}\xi\Big)\partial_z\mathbf{V}\Big]\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = 0. \tag{178}
$$

For the forth term on the right-hand side, we obtain with the inequalities of Hölder, Ladyshenzkaya and Young,

$$
\begin{aligned}
\Big|\int_\Omega (\mathbf{V}\cdot\nabla)\mathbf{v}\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\Big| &\leq c\|\nabla\mathbf{v}\|_{\mathbf{L}^6}\|\mathbf{V}\|_{\mathbf{L}^3}\|\partial_t\mathbf{V}\|_{\mathbf{L}^2} \\
&\leq c\|\nabla\mathbf{v}\|_{\mathbf{H}^1}\|\mathbf{V}\|_{\mathbf{H}^1}\|\partial_t\mathbf{V}\|_{\mathbf{L}^2} \leq \frac{c}{2\epsilon_1}\|\nabla\mathbf{v}\|_{\mathbf{H}^1}^2\|\mathbf{V}\|_{\mathbf{H}^1}^2 + \frac{\epsilon_1}{2}\|\partial_t\mathbf{V}\|_{\mathbf{L}^2}^2.
\end{aligned}
\tag{179}
$$

For the sixth term at the right-hand side of (177) we apply Lemma 2.2 and the Young inequality

$$
\begin{aligned}
\Big|\int_\Omega \Big[\Big(\int_{-h}^z \nabla\cdot\mathbf{V}(x,y,\xi,t)\,\mathrm{d}\xi\Big)\partial_z\mathbf{v}\Big]\cdot\partial_t\mathbf{V}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\Big| \\
\leq \|\partial_t\mathbf{V}\|_{\mathbf{L}^2}\|\mathbf{V}\|_{\mathbf{H}^1}^{1/2}\|\mathbf{V}\|_{\mathbf{H}^2}^{1/2}\|\partial_z\mathbf{V}\|_{\mathbf{L}^2}^{1/2}\|\partial_z\mathbf{V}\|_{\mathbf{H}^1}^{1/2} \\
\leq \frac{\epsilon_2}{2}\|\partial_t\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{1}{2\epsilon_2}\|\mathbf{V}\|_{\mathbf{H}^1}\|\mathbf{V}\|_{\mathbf{H}^2}\|\partial_z\mathbf{V}\|_{\mathbf{L}^2}\|\partial_z\mathbf{V}\|_{\mathbf{H}^1} \\
\leq \frac{\epsilon_2}{2}\|\partial_t\mathbf{V}\|_{\mathbf{L}^2}^2 + \frac{1}{2\epsilon_2}\|\mathbf{V}\|_{\mathbf{H}^1}^2\|\mathbf{V}\|_{\mathbf{H}^2}^2.
\end{aligned}
\tag{180}
$$

The remaining terms can be estimated with the inequality of Cauchy–Schwarz, and we obtain for (177) with the Young inequality

$$
\begin{aligned}
\|\partial_t\mathbf{V}\|_{\mathbf{L}^2}^2 &\leq c\big(\|\Delta\mathbf{V}\|_{\mathbf{L}^2}^2 + \|\partial_{zz}^2\mathbf{V}\|_{\mathbf{L}^2}^2 + \|\nabla\mathbf{v}\|_{\mathbf{H}^1}^2\|\mathbf{V}\|_{\mathbf{H}^1}^2 \\
&\quad + \|\mathbf{V}\|_{\mathbf{H}^1}^2\|\mathbf{V}\|_{\mathbf{H}^2}^2 + \|\nabla\Pi\|_{L^2}^2 + \|F_V\|_{\mathbf{L}^2}^2\big).
\end{aligned}
\tag{181}
$$

Integrating over $[0, T]$ and using the equations of state (32d) yields

$$\int_0^t ||\partial_t \mathbf{V}(s)||_{\mathbf{L}^2}^2 \mathrm{d}s \le c(\int_0^t ||\Delta \mathbf{V}(s)||_{\mathbf{L}^2}^2 + ||\partial_{zz}^2 \mathbf{V}(s)||_{\mathbf{L}^2}^2 + ||\nabla \mathbf{v}(s)||_{\mathbf{H}^1}^2 ||\mathbf{V}(s)||_{\mathbf{H}^1}^2$$
$$+ ||\mathbf{V}(s)||_{\mathbf{H}^1}^2 ||\mathbf{V}(s)||_{\mathbf{H}^2}^2 + ||\nabla \theta(s)||_{\mathbf{L}^2}^2 + ||F_{\mathbf{V}}(s)||_{\mathbf{L}^2}^2) \mathrm{d}s. \tag{182}$$

Using the regularity properties $\mathbf{v} \in L^\infty([0, T], \mathbf{H}^1(\Omega)) \cap L^2([0, T], \mathbf{H}^2(\Omega))$ and $\mathbf{V} \in L^\infty([0, T], \mathbf{H}^1(\Omega)) \cap L^2([0, T], \mathbf{H}^2(\Omega))$ we conclude that the right-hand side of (182) is finite. This proves $\partial_t \mathbf{V} \in L^2([0, T], \mathbf{L}^2(\Omega))$. With similar arguments follows that $\partial_t \Theta \in L^2([0, T], L^2(\Omega))$. From $\partial_t(V, \Theta) \in L^2([0, T], \mathcal{L}^2(\Omega))$ and $(V, \Theta) \in L^2([0, T], \mathcal{H}^2(\Omega))$ follows with the Aubins–Lions compactness lemma that $(V, \Theta) \in C([0, T], \mathcal{H}^1(\Omega))$. □

References

Agoshkov, V.I., Ipatova, V.M.: Solvability of the observation data assimilation problem in the three-dimensional model of ocean dynamics. Differ. Equ. **43**, 1088–110 (2007)

Azouani, A., Titi, E.S.: Feedback control of nonlinear dissipative systems by finite determining parameters—a reaction–diffusion paradigm. Equ. Control Theory **3**(4), 579–594 (2014)

Abergel, F., Temam, R.: On some control problems in fluid mechanics. Theor. Comput. Fluid Dyn. **1**, 303–325 (1990)

Bewley, T.R., Moin, P., Temam, R.: DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms. J. Fluid Mech. **447**, 179–225 (2001)

Bonavita, M., Lean, P., Holm, E.: Nonlinear effects in 4D-Var. Nonlinear Processes Geophys. **25**, 713–729 (2018)

Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, Berlin (2010)

Cao, C., Titi, E.S.: Global well-posedness of the three-dimensional viscous primitive equations of large scale ocean and atmosphere dynamics. Ann. Math. **166**, 245–267 (2007)

Cao, C., Titi, E.S.: Global Well-Posedness and finite-dimensional global attractor for a 3D planetary geostrophic viscous model. Commun. Pure Appl. Math. **56**, 198–233 (2003)

Chueshov, I.: A squeezing property and its applications to a description of long-time behaviour in the three-dimensional viscous primitive equations. Proc. R. Soc. Edinb. Sect. A **144**, 711–729 (2014)

Desamsetti, S., Dasari, H.P., Langodan, S., Titi, E.S., Knio, O., Hoteit, I.: Efficient dynamical downscaling of general circulation models using continuous data assimilation. Q. J. R. Soc. **145**, 3175–3194 (2019)

Evensen, G.: The Ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. **53**, 343–367 (2003)

Farhat, A., Lunasin, E., Titi, E.S.: On the Charney conjecture of data assimilation employing temperature measurements alone: the paradigm of 3D planetary geostrophic model. Math. Clim. Weather Forecast. **2**, 61–74 (2016)

Foias, C., Mondaini, S., Titi, E.S.: A Discrete Data Assimilation Scheme for the Solutions of the 2D Navier–Stokes Equations and their Statistics. SIAM J. Appl. Dyn. Syst. **15**, 2109–2142 (2016)

Forget, G., Campin, J.-M., Heimbach, P., Hill, C.N., Ponte, R.M., Wunsch, C.: ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. Geosci. Model Dev. **8**, 3071–3104 (2015)

Fursikov, A.V.: Optimal Control of Distributed Systems. American Mathematical Society, Providence (1999)

Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S., Morneau, J.: Extension of 3DVAR to 4DVAR: implementation of 4DVAR at the meteorological service of Canada. Mon. Weather Rev. **135**, 2339–2354 (2007)

Guillén-González, F., Masmoudi, N., Rodríguez-Bellido, M.A.: Anisotropic estimates and strong solutions of the primitive equations. Differ. Integral Equ. **14**, 1381–1408 (2001)

Hieber, M., Kashiwabara, T.: Global strong well-posedness of the three dimensional primitive equations in L^p-spaces. Arch. Ration. Mech. Anal. **221**, 1077–1115 (2016)

Gibbon, J.D., Holm, D.D.: Enstrophy bounds and the range of space-time scales in the hydrostatic primitive equations. Phys. Rev. E **87**, 031001 (2013)

Gibbon, J.D., Holm, D.D.: Extreme events in solutions of hydrostatic and non-hydrostatic climate models. Philos. Trans. R. Soc. A **369**, 1156–1179 (2011)

Hoteit, I., Cornuelle, B., Köhl, A., Stammer, D.: Treating strong adjoint sensitivities in tropical eddy-permitting variational data assimilation. Q. J. Roy. Soc. **131**, 3659–3682 (2005)

Houtekamer, P.L., Zhang, F.: Review of the ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. **144**, 4489–4532 (2016)

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., Raynaud, L.: Ensemble of data assimilations at ECMWF. ECMWF Technical memorandum 636. https://www.ecmwf.int/en/elibrary/10125-ensemble-data-assimilations-ecmwf

Ju, N.: The global attractor for the solutions to the 3D viscous primitive equations. Discrete Contin. Dyn. Syst. **17**, 711–729 (2007)

Kalnay, E.: Atmosheric Modeling, Data Assimilation and Predictability. Cambridge University Press, Cambridge (2003)

Kobelkov, G.M.: Existence of a solution "in the large" for ocean dynamics equations. J. Math. Fuid Mech. **10**, 1–23 (2007)

Köhl, A., Willebrand, J.: An adjoint method for the assimilation of statistical characteristics into eddy-resolving ocean models. Tellus **54**, 406–425 (2002)

Korn, P.: A regularity-aware algorithm for variational data assimilation of an idealized coupled atmosphere-ocean model. J. Sci. Comput. **79**, 748–786 (2019)

Korn, P.: Data assimilation for the Navier–Stokes-α equations. Physica D **238**, 1957–1974 (2009)

Kukavica, I., Pei, Y., Rusin, W., Ziane, M.: Primitive equations with continuous initial data. Nonlinearity **27**, 1–21 (2014)

Kukavica, I., Ziane, M.: On the regularity of the primitive equations of the ocean. Nonlinearity **20**, 2739–2753 (2007)

Lions, J.L., Temam, R., Wang, S.: On the equation of the large scale ocean. Nonlinearity **5**, 1007–1053 (1992)

Li, J., Titi, E.S.: The primitive equations as the small aspect ratio limit of the Navier-Stokes equations: Rigorous justification of the hydrostatic approximation. J. Math Pures et Appl., **124**, 30–58 (2019)

Li, J., Titi, E.S.: Recent advances concerning certain class of geophysical flows. In: Giga, Y., Novotny, A. (eds.) Handbook of Mathematical Analysis in Mechanics of Viscous Fluids. Springer, Berlin (2018)

Lorenc, A.: Modelling of error covariances by 4D-Var data assimilation. Q. J. R. Meteorol. Soc. **129**, 3167–3182 (2003)

Marchuk, G.I., Agoshkov, V.I., Shutyaev, V.P.: Adjoint Equations and Perturbation Algorithms in Nonlinear Problems. CRC Press, Boca Raton (1996)

Necas, J.: Direct Methods in the Theory of Elliptic Equations. Springer, Berlin (2012)

Pei, W.: Continuous data assimilation for the 3D primitive equations of the ocean. Commun. Pure Appl Anal. **19**, 643–661 (2019)

Petcu, M., Temam, R., Ziane, M.: Some mathematical problems in geophysical fluid dynamics. In: Friedlander, S., Serre, D. (eds.) Handbook of Mathematical Fluid Dynamics, vol. 14, pp. 535–657. North Holland, Amsterdam (2009)

Reich, S., Cotter, C.: Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge University Press, Cambridge (2015)

Schillings, C., Stuart, A.M.: Analysis of the ensemble Kalman filter for inverse problems. SIAM J. Numer. Anal. **55**, 1264–1290 (2017)

Schillings, C., Stuart, A.M.: Convergence analysis of ensemble Kalman inversion: the linear, noisy case. Appl. Anal. **97**, 107–123 (2018)

Shutyaev, V.P.: Solvability of the data assimilation problem in the scale of Hilbert spaces for quasilinear singularly perturbed evolutionary problems. Russ. J. Numer. Anal. Math. Model. **12**, 53–66 (1997)

Medjo, T.Tachim, Temam, R., Ziane, M.: Optimal and robust control of fluid flow: some theoretical and computational aspects. Appl. Mech. Rev. **61**, 1–23 (2008)

Weaver, A., Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation. Q. J. R. Meteorol. Soc. **127**, 1815–1846 (2001)

Wunsch, C.: The Ocean Circulation Inverse Problem. Cambridge University Press, Cambridge (1997)