

Unifying Physical Backpropagation

Cyrill Bösch^{*1}, Yiğithan Gediz², Clara C. Wanjura³, and Hakan E. Türeci²

¹Department of Computer Science, Princeton University, Princeton, NJ 08540, USA

²Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540, USA

³Max Planck Institute for the Science of Light, Staudtstraße 2, 91058 Erlangen, Germany

September 22, 2026

Abstract

Physical computing systems exploit device dynamics for computation, but their gradient-based optimization is challenging: backpropagation through a digital twin suffers from a model–reality gap. On-device gradient computation could resolve this issue, and a handful of theoretical and experimental studies have proposed ways to achieve it. Yet a unifying theory identifying when a physical system can compute the gradient of its own performance has been missing. Here we develop such a unification based on the adjoint method: we identify sufficient conditions under which the adjoint field required for formally exact gradients can be generated on the same hardware that performs the computation. Linear and nonlinear systems obey fundamentally different conditions: for linear systems, damping or gain is admissible provided reciprocity is preserved. For nonlinear trajectory systems, the sufficient conditions are reciprocity of the linearized system and the existence of a time-reversal mirror. Algorithmically, the nonlinear case requires infinitesimal nudging, whereas linear systems admit a finite-amplitude experiment. We recover (quantum) Equilibrium Propagation, Hamiltonian echo backpropagation, fully forward mode training and in situ gradient methods in integrated-photonics and free-space-optical systems. Finally, we show that reciprocity is a special case of more general intertwining conditions. For linear systems, these permit exact on-device gradients in a class of non-Hermitian, non-reciprocal systems. For nonlinear trajectories, they combine with generalized time-reversal mirrors to cover, e.g., $\mathcal{PT}$ -symmetric equations. The framework also includes time-dependent parameters and Onsager-reciprocal dynamics, providing a unified basis for formally exact physical learning.

Contents

1	Introduction	2
2	Introducing the adjoint method	3
3	What is computed physically, under what conditions and how: an overview	6
4	Local learning rules and “knowing the model” vs. “knowing the parameters”	7
5	Second-order real-valued systems	9
6	First-order real-valued systems	13

^{*}cyrillboesch@gmail.com

7	Schrödinger-type complex systems	15
8	Stationary problems and (quantum) Equilibrium Propagation	18
9	General intertwiners and twisted reciprocity	23
10	Conclusions	28
	Appendix	36

1 Introduction

Physical computing uses the controlled dynamics or steady-state response of a material or device for information processing [1–3]. Inputs may be encoded in initial or boundary conditions, external drives, or programmable parameters, and outputs are measured from selected physical degrees of freedom. Representative platforms include mechanical and elastic networks [4], analog electrical circuits [5], spintronic and memristive devices [6, 7], integrated photonic circuits [8], nonlinear multimode fibers [9], free-space optical systems [10–12] and programmable waveguide systems [13, 14]. These platforms span transient, fixed-point, and frequency-domain computations.

A common special case is Physical Reservoir Computing [15], in which the internal physical dynamics is fixed and only a linear readout layer is trained. Here, instead, we address the optimization of parameters that enter the physical dynamics itself, which remains a major challenge [16]. Training only a digital twin [17] is vulnerable to a model–reality gap. Physics-aware training, in which the forward pass is performed in the physical device but differentiation is performed using a digital model, can reduce this gap, but still requires a differentiable surrogate for the backward pass and communication between the physical and digital domains [18, 19]. Over the past decade, *in situ* (equivalently, on-device) training methods, in which the physical system participates in both the forward computation and the generation of the learning signal, have emerged as a promising alternative [20]. These approaches range from heuristic or approximate rules to methods that compute formally exact gradients. Approximate rules can be simple and local, but their updates need not coincide with the true gradient, and their convergence and performance can therefore be system dependent [16]. This motivates formally exact physical-gradient constructions, which have been developed for free-space optics [21, 22], integrated photonics [23], mechanics [24], resistive electrical networks [25], and Hamiltonian systems [26], among other platforms.

In this article, we unify these formally exact physical backpropagation methods using the adjoint framework of PDE-constrained optimization [27, 28] and extend them to broader classes of physical systems. The main contribution of this work is twofold. First, we show that seemingly disparate physical backpropagation algorithms are instances of a single adjoint construction: this common formulation lets us compare and extend these algorithms. Second, we identify sufficient conditions under which the required adjoint field, and hence the cost function gradient, can be obtained from a physical experiment that runs in the *same* system. The only allowed modifications between the forward and adjoint experiments are changes of initial conditions, external forces, and reversal of explicit time dependences.

We show that linear and nonlinear systems give rise to different physical experiments and different conditions. Linear reciprocal systems admit direct, finite-amplitude adjoint experiments, and damping or gain is compatible with exact physical gradients as long as reciprocity is preserved. For example, trajectory gradients can be obtained physically even in linear reciprocal overdamped systems [20], whereas the corresponding nonlinear trajectory construction is obstructed. Under suitable conditions, Equilibrium Propagation [29] emerges for fixed-point engineering.

Nonlinear trajectory systems are more restrictive: they require infinitesimal nudging, a particular form of time-reversal symmetry and reciprocity of the linearized dynamics. The latter two are independent conditions. In the nonlinear regime, ordinary damping or gain is generically obstructive, and even the

ability to experimentally swap damping and gain does not in general restore a same-hardware adjoint construction.

The adjoint framework recovers a broad range of formally exact physical backpropagation algorithms: Equilibrium Propagation [29], the exact limit of complex-valued Near-Equilibrium Propagation [30], Quantum Equilibrium Propagation [31–33], Hamiltonian echo backpropagation and its recurrent variants [26, 34], early reciprocity-based recurrent learning [20], and physical gradient algorithms from integrated photonics [23, 35], free-space and diffractive optics [21, 22], linear wave scattering [36, 37], static and topological mechanics [24, 38], and the projector-based analytical gradient for passive resistor networks [25]. For optical schemes with approximate gradients through the nonlinear subsystems [39–41], the framework recovers only the exact backpropagation through the linear layers. We then show that ordinary reciprocity can be generalized to an intertwining condition, in which the forward operator and its transpose are related by a constant invertible transformation. This construction yields exact physical gradients in a class of non-Hermitian and non-reciprocal systems. In the linear case, we call such systems “twisted reciprocal” and illustrate them with a Hatano–Nelson chain [42]. Fully Forward Mode (FFM) training [43] is a special case of this construction. For nonlinear systems, we combine intertwining with a suitable time-reversal mirror and give examples involving nonlinear Hatano–Nelson chains and $\mathcal{PT}$ -symmetric dynamics. Onsager reciprocity provides another special case, with the intertwining transformation supplied by a physical mobility (Sec. 9.3). We further establish sufficient conditions for physical backpropagation in non-autonomous systems, making the temporal modulation of internal forces a trainable resource.

Disclaimer. In this article we do not cover coupled [5, 44–48] and heuristic [49–51] physical learning rules. We also do not discuss forward-forward and related local-learning schemes for physical systems, which forgo gradients of a global loss function in favor of layer-local objectives [52–54]. Furthermore, we do not cover probabilistic physical computing machines. Boltzmann machines [55] and score-based generative models [56] have been shown to give rise to exact (unbiased) local learning rules, although these do not derive from the adjoint method. We consider a setup where the gradient is calculated from physically observed forward and adjoint fields. We are agnostic to how the parameter updates are then applied and assume for simplicity that an external teacher does so. This means we do not cover the problem of designing a physical system that adapts by itself [5, 26, 39, 46].

Structure. In Sec. 2 we formally define the optimization problem and introduce the adjoint method using a second-order real-valued system. In Sec. 3 we precisely define which part of the gradient computation is performed by the physical system and provide a summary of our results across various systems. In Sec. 4 we introduce the notion of local learning rules and discuss the difference between knowing the model and knowing the parameters. In Sections 5–8 we present the main results: sufficiency conditions for obtaining the adjoint fields physically in second- (Sec. 5) and first-order overdamped (Sec. 6) real-valued systems, in Schrödinger-type complex-valued systems (Sec. 7), and in stationary problems (Sec. 8), including Equilibrium Propagation, Quantum Equilibrium Propagation (Sec. 8.3), and Helmholtz-type constructions. At the end of each section we identify which existing physical learning frameworks are recovered by the adjoint formalism. Section 9 treats general intertwiners and time-reversal mirrors, including fully forward mode training and twisted-reciprocal Hatano–Nelson chains. Section 9.3 discusses Onsager reciprocity within this construction.

A self-contained appendix derives and discusses each case in more detail. We also provide a notebook for each section, with an implementation of the corresponding physical adjoint algorithm [57].

2 Introducing the adjoint method

We consider N -dimensional discretized systems governed by ODEs, though the theory also applies to fields governed by PDEs. If the ODE resulted from discretizing a continuous PDE through, e.g., the finite element method, then, in the language of PDE-constrained optimization, this would correspond to the principle of

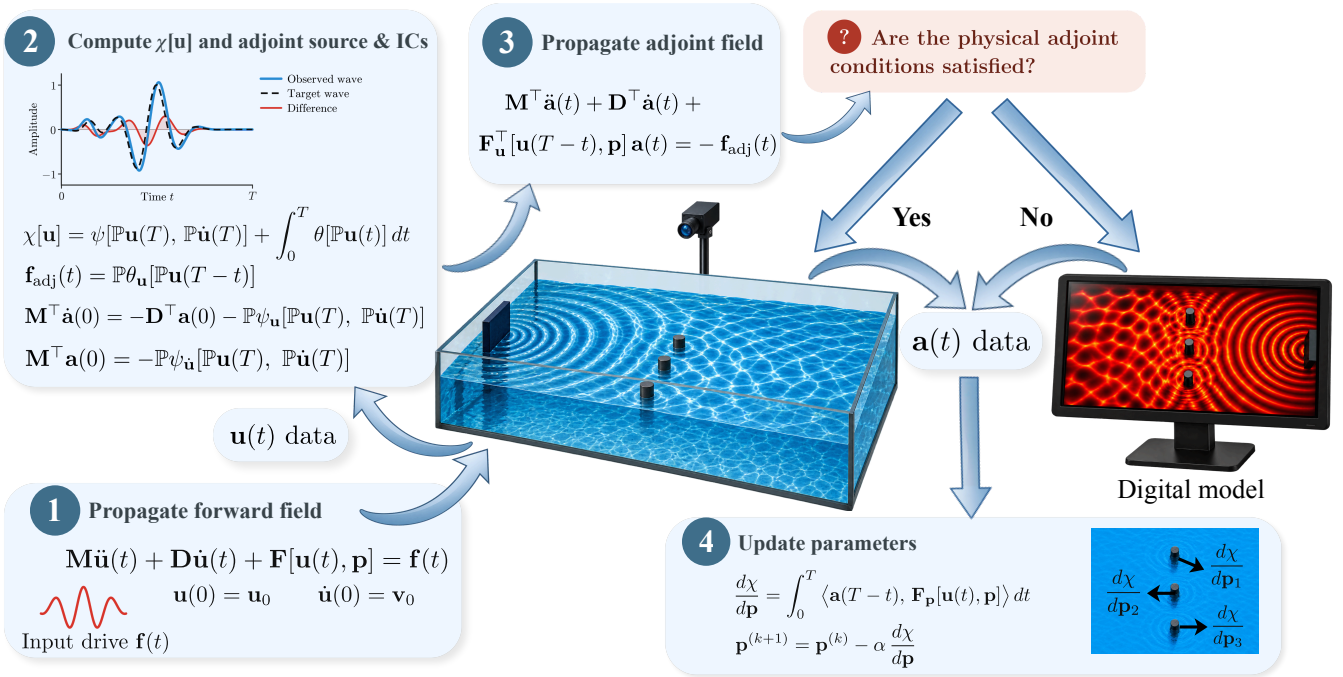


Figure 1: Conceptual overview of optimizing a physical system with the adjoint method. (1) The physical state $\mathbf{u}(t)$ evolves under the forward dynamics $\mathbf{M}\ddot{\mathbf{u}} + \mathbf{D}\dot{\mathbf{u}} + \mathbf{F}[\mathbf{u}, \mathbf{p}] = \mathbf{f}(t)$, with mass matrix $\mathbf{M}$, damping matrix $\mathbf{D}$, internal forces $\mathbf{F}$ carrying the trainable parameters $\mathbf{p}$, external drive $\mathbf{f}$, and initial conditions $\mathbf{u}(0) = \mathbf{u}_0$, $\dot{\mathbf{u}}(0) = \mathbf{v}_0$. (2) The recorded trajectory and terminal state determine the cost $\chi[\mathbf{u}] = \psi[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)] + \int_0^T \theta[\mathbb{P}\mathbf{u}(t)] dt$, where $\mathbb{P}$ projects onto the measured output degrees of freedom and ψ and θ are the terminal and trajectory losses. Evaluating the derivatives of the losses provides the drive and initial conditions (ICs) for the adjoint field $\mathbf{a}(t)$. (3) The adjoint dynamics involve the transposed mass and damping matrices and the linearized internal forces $\mathbf{F}_{\mathbf{u}}^T$, evaluated along the time-reversed forward trajectory $\mathbf{u}(T-t)$. They therefore generally differ from the forward dynamics. A physical forward pass combined with a backward pass through a differentiable digital model is known as physics-aware training [18, 19]. Here we ask under what conditions the same physical hardware can generate the adjoint field, without a digital model for the backward pass. Finally, (4) the gradient $d\chi/d\mathbf{p}$ is evaluated from overlap integrals of the recorded forward and adjoint fields, and the physical parameters are updated.

“discretize-then-optimize” [28]. We denote the physical state by $\mathbf{u}(t) \in \mathbb{R}^N$ and the trainable parameters by $\mathbf{p} \in \mathbb{R}^M$. Inputs may be encoded in initial conditions, external drives [58], or directly in parameters. The latter endows linear systems with nonlinear computing capabilities [36, 59, 60]. Suppressing explicit encoders, decoders, and batch averages, we write the cost as

$$\chi[\mathbf{u}] = \int_0^T \theta[\mathbb{P}\mathbf{u}(t)] dt + \psi[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)], \quad (1)$$

where $\mathbb{P} = \mathbb{P}^T = \mathbb{P}^2$ projects onto the measured or output degrees of freedom (DOFs) which encode the solution to the computation. The first term is a trajectory loss and the second a final-state loss.

We now introduce the adjoint method [27, 61–68] that we will use throughout the text. The constructions used here are rooted in PDE-constrained optimization, where an objective is minimized subject to a governing equation enforced as a constraint through Lagrange multipliers [69]. A full, self-contained derivation is given in Appendix A.

For concreteness, consider a second-order real-valued system

$$\mathbf{M}(\mathbf{p}_M)\ddot{\mathbf{u}}(t) + \mathbf{D}(\mathbf{p}_D)\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, \quad \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}), \quad \dot{\mathbf{u}}(0) = \mathbf{v}_0(\mathbf{p}_{v_0}). \quad (2)$$

Here $\mathbf{M}$ is the invertible mass matrix, $\mathbf{D}$ the damping or gain matrix, $\mathbf{F}$ the internal force, which carries the constitutive law of the system and is in general nonlinear in $\mathbf{u}$ and explicitly time dependent, and $\mathbf{f}$ the external drive; $\mathbf{u}_0$ and $\mathbf{v}_0$ are the initial state and its rate of change. We use mechanical language throughout, but the same three operators appear as the inertial, dissipative, and restoring terms of electrical, optical, and other physical platforms. Each operator carries its own block of trainable parameters, collected into $\mathbf{p} = (\mathbf{p}_M, \mathbf{p}_D, \mathbf{p}_F, \mathbf{p}_f, \mathbf{p}_{u_0}, \mathbf{p}_{v_0})$. The constrained optimization problem is

$$\min_{\mathbf{p}} \chi[\mathbf{u}] \quad \text{subject to} \quad (2).$$

The adjoint method replaces the differentiation of the full trajectory $\mathbf{u}(t; \mathbf{p})$ with respect to every parameter by a single adjoint solve. Written in forward time, the adjoint field $\mathbf{a}(t)$ satisfies (see Appendix A)

$$\mathbf{M}^T \ddot{\mathbf{a}}(t) + \mathbf{D}^T \dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T-t), \mathbf{p}_F, T-t] \mathbf{a}(t) + \mathbf{f}_{\text{adj}}(t) = \mathbf{0}, \quad (4)$$

where $(\cdot)_{\mathbf{x}}$ denotes a partial derivative, i.e. $\mathbf{F}_{\mathbf{u}} := \partial \mathbf{F} / \partial \mathbf{u}$ (in the parameters, $\mathbf{p}_X$, the subscripts are labels). The adjoint source is given by

$$\mathbf{f}_{\text{adj}}(t) = \mathbb{P} \theta_{\mathbf{u}}[\mathbb{P} \mathbf{u}(T-t)] \quad (5)$$

and the adjoint initial conditions are

$$\mathbf{M}^T \mathbf{a}(0) = -\mathbb{P} \psi_{\dot{\mathbf{u}}}[\mathbb{P} \mathbf{u}(T), \mathbb{P} \dot{\mathbf{u}}(T)], \quad (6)$$

$$\mathbf{M}^T \dot{\mathbf{a}}(0) = -\mathbf{D}^T \mathbf{a}(0) - \mathbb{P} \psi_{\mathbf{u}}[\mathbb{P} \mathbf{u}(T), \mathbb{P} \dot{\mathbf{u}}(T)]. \quad (7)$$

The derivative of the trajectory loss, $\mathbb{P} \theta_{\mathbf{u}}$, is therefore a forcing term in the adjoint experiment. The derivatives of the final-state loss, $\mathbb{P} \psi_{\mathbf{u}}$ and $\mathbb{P} \psi_{\dot{\mathbf{u}}}$, set the adjoint initial displacement and velocity. If there is no trajectory loss, the adjoint is unforced; if there is no terminal loss, the adjoint starts from zero.

Once $\mathbf{a}(t)$ is known, the gradient is obtained from overlap integrals (see Appendix A.9)

$$\frac{d\chi}{d\mathbf{p}_M} = \int_0^T \langle \mathbf{a}(T-t), \mathbf{M}_{\mathbf{p}_M} \ddot{\mathbf{u}}(t) \rangle dt, \quad (8)$$

$$\frac{d\chi}{d\mathbf{p}_D} = \int_0^T \langle \mathbf{a}(T-t), \mathbf{D}_{\mathbf{p}_D} \dot{\mathbf{u}}(t) \rangle dt, \quad (9)$$

$$\frac{d\chi}{d\mathbf{p}_F} = \int_0^T \langle \mathbf{a}(T-t), \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] \rangle dt, \quad (10)$$

$$\frac{d\chi}{d\mathbf{p}_f} = - \int_0^T \langle \mathbf{a}(T-t), \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}_f) \rangle dt, \quad (11)$$

$$\frac{d\chi}{d\mathbf{p}_{u_0}} = - \langle [\mathbf{M}^T \dot{\mathbf{a}}(T) + \mathbf{D}^T \mathbf{a}(T)], \mathbf{u}_{0, \mathbf{p}_{u_0}} \rangle, \quad (12)$$

$$\frac{d\chi}{d\mathbf{p}_{v_0}} = - \langle \mathbf{M}^T \mathbf{a}(T), \mathbf{v}_{0, \mathbf{p}_{v_0}} \rangle, \quad (13)$$

where $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ is the Euclidean inner product. Thus the gradient with respect to dynamical parameters is an overlap between the adjoint field and the parameter sensitivity of the respective forward operators. The gradient with respect to initial-condition parameters is an endpoint overlap. In particular, if the initial displacement and velocity themselves are trainable, then

$$\frac{d\chi}{d\mathbf{u}_0} = -\mathbf{M}^T \dot{\mathbf{a}}(T) - \mathbf{D}^T \mathbf{a}(T), \quad \frac{d\chi}{d\mathbf{v}_0} = -\mathbf{M}^T \mathbf{a}(T). \quad (14)$$

Interpretation. The overlap integrals show that the time-reversed adjoint field $\mathbf{a}(T - t)$ encodes the sensitivity of the loss with respect to a force applied at time t during the forward evolution. All parameters act in this way. Writing the state equation as $\mathbf{M}\ddot{\mathbf{u}} + \mathbf{D}\dot{\mathbf{u}} + \mathbf{F} = \mathbf{f}$, a parameter change $\delta\mathbf{p}$ perturbs the net force acting on the system along the forward trajectory by

$$\delta\mathbf{f}(t) = \mathbf{f}_{\mathbf{p}_f} \delta\mathbf{p}_f - \mathbf{M}_{\mathbf{p}_M} \ddot{\mathbf{u}}(t) \delta\mathbf{p}_M - \mathbf{D}_{\mathbf{p}_D} \dot{\mathbf{u}}(t) \delta\mathbf{p}_D - \mathbf{F}_{\mathbf{p}_F} [\mathbf{u}(t), \mathbf{p}_F, t] \delta\mathbf{p}_F, \quad (15)$$

where the internal contributions enter with the opposite sign because they act against the drive. The trajectory gradients (8)–(11) then collapse into the single statement $\delta\chi = -\int_0^T \langle \mathbf{a}(T - t), \delta\mathbf{f}(t) \rangle dt$. If the force change aligns with the adjoint field, the cost therefore decreases at the largest rate per unit of forcing, while a force orthogonal to it leaves the cost unchanged. The endpoint overlaps (12) and (13) have the same interpretation for the initial conditions.

The arrow of time. As discussed in Appendix A.10, the adjoint state as it emerges from the derivation evolves backward in time. The equations above already incorporate the change of variable $t \mapsto T - t$ that recasts this evolution in forward time; this will become crucial for the physical realization of the adjoint field in Sec. 5.

Algorithm. The algorithm is summarized in Fig. 1. One first solves the forward problem for the forward field $\mathbf{u}(t)$, either digitally or, in physics-aware training, on the physical system itself. Next, the adjoint problem, which in general differs from the forward problem through the transposes and the linearization, is solved digitally for $\mathbf{a}(t)$. The gradient then follows from the overlap integrals in (8)–(13), and the parameters are finally updated according to a chosen optimization algorithm.

A brief note on computational complexity. The advantage of the adjoint method in digital optimization is that the cost of computing the gradient is essentially independent of the number of parameters [28, 61, 62]. Finite-difference methods and forward sensitivity analysis require a separate solve for every entry of $\mathbf{p}$, so their computational cost grows linearly with the parameter count $\dim(\mathbf{p})$. The adjoint method, by contrast, requires only two trajectory solves regardless of how many parameters are trained: one forward solve for $\mathbf{u}(t)$ and one adjoint solve for $\mathbf{a}(t)$. All gradient components are then read off from the overlap integrals (8)–(13), which are cheap because the only ingredients they require are the parameter sensitivities of the forward operators, such as $\mathbf{M}_{\mathbf{p}_M}$, $\mathbf{D}_{\mathbf{p}_D}$, and $\mathbf{F}_{\mathbf{p}_F}$, which are typically sparse, structured, and available in closed form from the model itself (see Sec. 4).

How to test the gradient? An adjoint gradient implementation can be checked with a Taylor test. We detail this procedure in Appendix J. We provide a notebook for each section, with an implementation of the corresponding physical adjoint algorithm and a Taylor test that numerically verifies the computed gradients [57].

Forward and adjoint systems are different. The adjoint construction is exact and fully general: it applies to any physical system, linear or nonlinear, conservative or dissipative. The adjoint problem (4) involves transposes and derivatives of the internal forces, and is hence in general *different* from the physical system that evolves with the forward state equation (2).

3 What is computed physically, under what conditions and how: an overview

In this article we ask the following question (see Fig. 1):

Under what conditions, and by which experimental protocol, can the adjoint field be obtained in the *same physical system* that performs the forward computation?

By the same system we mean the same hardware and trainable parameters; we may only change initial conditions and applied sources, and reverse explicit time dependences.

With one exception in the readout, all constructions below have the same form: run the forward dynamics and record $\mathbf{u}(t)$; construct adjoint sources (Eq. (5)) and initial conditions (Eqs. (6), (7)) from the loss derivatives; obtain the adjoint field $\mathbf{a}(t)$ from a second physical experiment; then evaluate the gradient by an overlap between the measured forward/adjoint fields and known parameter sensitivities (Eqs. (8)–(13)). Quantum Equilibrium Propagation (Sec. 8.3) obtains the gradient directly from changes in operator expectation values without reading out the adjoint field. The distinction between linear and nonlinear systems is the central organizing principle of the article. Across all systems we consider, the conditions and algorithms can be summarized as follows:

Sufficient physical adjoint conditions and constructions.

- **Linear systems:** Reciprocity is sufficient; Hermiticity is not required, so loss and gain are permitted. The adjoint field is obtained by a direct finite-amplitude run of the same physical system with adjoint sources and initial conditions and any explicit time dependence played in reverse.
- **Nonlinear trajectory systems:** The forward evolution must be recoverable by a time-reversal operation solely on the final state and the linearized dynamics around the reversed trajectory must be reciprocal. The adjoint field is then obtained as the infinitesimal response of a nudged time-reversed trajectory. Explicit time dependences are played in reverse.
- **Stationary linear problems:** In linear reciprocal fixed-point or Helmholtz-type systems the adjoint field can be obtained directly with a single finite-amplitude experiment.
- **Stationary nonlinear problems:** In nonlinear stationary systems, the exact adjoint is obtained in the infinitesimal limit from the difference between a free solution and a nudged solution, provided the system relaxes to the steady state and the linearized operator there is reciprocal. This is Equilibrium Propagation [29].

Complex-valued systems require separate conditions on the two Wirtinger tangent operators. Sections 5–8 develop these basic constructions. Section 9 extends them through general intertwiners and time-reversal mirrors; Onsager reciprocity is the mobility-weighted special case discussed in Sec. 9.3.

Table 1 organizes the cases treated in the remainder of the article along these lines. For each case, it records whether the adjoint field is available on the device and which experiment or proposal in the literature realizes it.

4 Local learning rules and “knowing the model” vs. “knowing the parameters”

If the adjoint field can be observed physically and the parameter sensitivity is local, the remaining overlap gives a local learning rule. Consider two coupled Duffing oscillators with internal force

$$\mathbf{F}(\mathbf{u}, \mathbf{p}) = \begin{bmatrix} k_1 u_1 + \beta_1 u_1^3 + c(u_1 - u_2) \\ k_2 u_2 + \beta_2 u_2^3 + c(u_2 - u_1) \end{bmatrix}, \quad \mathbf{p} = (k_1, k_2, \beta_1, \beta_2, c). \quad (16)$$

Table 1: The cases treated in this article. The trajectory constructions also cover non-autonomous dynamics with programmable internal forces: replay their schedules in reverse, with the stated conditions holding throughout the run. “On device” denotes a construction under the stated conditions, allowing known invertible field transformations. A $\blacklozenge$ marks an experimental realization; a $\star$ marks a case without a known realization. Prescribed trajectory compensation is excluded; see the [second-order](#) and [first-order](#) discussions.

	System	Sufficient condition	Adjoint construction	Sec.	On device	Representative realization
Linear systems	Reciprocal					
	Second order, real	$\mathbf{M}^T = \mathbf{M}$ and $\mathbf{D}^T = \mathbf{D}$, both constant in time, and $\mathbf{K}^T = \mathbf{K}$; damping is no obstacle	Forward drive and initial conditions (ICs) swapped for adjoint source and ICs and applied at full amplitude	5.1	yes	$\blacklozenge$ acoustic proof of concept [20]; $\blacklozenge$ coupled RLC network [71], written there in first-order form
	First order, real	$\mathbf{K}^T = \mathbf{K}$	As above	6.1	yes	Reciprocal dynamical construction [20]
	Schrödinger type	$\mathbf{K}^T = \mathbf{K}$; reciprocity and not Hermiticity, so gain and loss are admissible	As above; what propagates is the conjugate adjoint $\mathbf{c} = \bar{\mathbf{a}}$	7.1	yes	In situ optical backpropagation [21]; linear layers of holographic networks [39]
	Stationary, Helmholtz	$\mathbf{K}^T = \mathbf{K}$	Forward drive swapped for the adjoint source; applied at full amplitude	8.1	yes	Photonic meshes [35]; $\blacklozenge$ Pai et al. [23]. Wave scattering [36]; $\blacklozenge$ Guillaumon et al. [37]
	Intertwined reciprocity					
	General intertwiner	Known, time-independent, invertible $\mathbf{S}$ with $\mathbf{S}\mathbf{K}^T\mathbf{S}^{-1} = \mathbf{K}$; the same relation for $\mathbf{M}, \mathbf{D}$ when present	One run with $\mathbf{S}$ -weighted sources and ICs; recover $\mathbf{c} = \bar{\mathbf{a}} = \mathbf{S}^{-1}\mathbf{d}$	9	yes	Weighted stationary transpose access in resistor networks [25]
Nonlinear systems	Hardware symmetry	$\mathbf{K}^T = \mathbf{K}$ and $\mathbf{S}\mathbf{K}\mathbf{S}^{-1} = \mathbf{K}$ with $\mathbf{S}\mathbf{P}\mathbf{S}^{-1} = \mathbb{P}^{\text{in}}$	The error signal enters where the data does and propagates forward; recover $\mathbf{c} = \bar{\mathbf{a}} = \mathbf{S}^{-1}\mathbf{d}$	9	yes	$\blacklozenge$ Fully forward mode training [43], in static optics
	Twisted reciprocity	Orthogonal $\mathbf{S}$ with $\mathbf{S}\mathbf{K}\mathbf{S}^{-1} = \mathbf{K}^T$; e.g. a mirror-even Hatano–Nelson chain	One run with relocated sources; recover $\mathbf{c} = \bar{\mathbf{a}} = \mathbf{S}^{-1}\mathbf{d}$	9	yes	$\star$ none
	Trajectories					
	Second order, real, undamped	$\mathbf{D} = \mathbf{0}$, so a time-reversal mirror exists, and $\mathbf{M}^T = \mathbf{M}$ with $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ along the reversed trajectory	Infinitesimal response of a nudged time-reversed trajectory	5.2	yes	Hamiltonian echo backpropagation [26]; recurrent Hamiltonian echo learning [34]
	Second order, real, damped	None: time reversal turns damping into gain but the adjoint requires damping	Obstructed	5.2	no	Contrastive trajectory rule [72], approximate
	First order, real, overdamped	None: reversal requires $\mathbf{F} \mapsto -\mathbf{F}$, the adjoint $+\mathbf{F}_{\mathbf{u}}^T$	Obstructed; only the steady state survives	6.2	no	Rayleighian trajectory rule [73], approximate
	Schrödinger	Conjugation is a TRM; $\mathbf{F}_{\mathbf{u}}$ is Hermitian and $\mathbf{F}_{\bar{\mathbf{u}}}$ symmetric along the trajectory	Infinitesimal response of a nudged conjugation-reversed trajectory gives $\mathbf{c} = \bar{\mathbf{a}}$	7.2	yes	Hamiltonian echo backpropagation in nonlinear waves [26]
	Stationary problems					
	Real fixed point (Equilibrium Propagation)	A stable fixed point, and $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ there	Difference between the free and the nudged equilibrium	8.2	yes	Equilibrium Propagation [29]; $\blacklozenge$ Ising-machine implementation [74]
	Complex steady state (Complex-valued EP)	A stable steady state. For $\mathbf{a}$: $\mathbf{F}_{\mathbf{u}}$ Hermitian and $\mathbf{F}_{\bar{\mathbf{u}}}$ symmetric. For $\mathbf{c} = \bar{\mathbf{a}}$: $\mathbf{F}_{\mathbf{u}}$ symmetric and $\mathbf{F}_{\bar{\mathbf{u}}}$ Hermitian	Infinitesimal response; use the adjoint-source nudge for $\mathbf{a}$ and its complex conjugate for $\mathbf{c}$	8.2	yes	Near-Equilibrium Propagation (NEP) recovers $\mathbf{a}$ in the exact limit [30]; $\blacklozenge$ experiment recovering $\mathbf{c} = \bar{\mathbf{a}}$ with one nonlinear element [75]
Nonlinear systems	Quantum ground state (QEP)	Hermitian Hamiltonian with a nondegenerate, gapped ground state; tangent reciprocity holds after normalization and phase fixing	Adjoint source enters as a Hamiltonian perturbation; measure operator expectations in the free and nudged states without recovering the adjoint field	8.3	yes	Quantum Equilibrium Propagation [31–33]
	Non-reciprocal tangent	Direct reciprocity fails, so the unweighted nudged response generally does not solve the adjoint equation	Unweighted nudge obstructed	8.2	no	Approximate extensions [76, 77]; [78] modifies the learning dynamics
	Tangent intertwiners					
	Trajectories: second order and Schrödinger	Known, constant invertible maps $\mathcal{Q}, \mathbf{S}$: $\mathcal{Q}$ generates a TRM, and $\mathbf{S}$ matches the adjoint to the pulled-back tangent, including inertial and velocity terms	Infinitesimal response with transformed sources and ICs; undo $\mathcal{Q}\mathbf{S}$. For complex fields the TRM is $\mathcal{Q}\mathcal{K}$	9.2, D.1, G	yes	$\star$ no known realization in this generality; related $\mathcal{PT}$ -assisted Schrödinger protocol with gain–loss compensation [79]
	Stationary states	A stable state with invertible full tangent; a known invertible $\mathbf{S}$ maps the adjoint tangent onto the physical tangent	Infinitesimal response to a transformed nudge on the same solution branch; undo $\mathbf{S}$	9.3	yes	EP derivation with Onsager mobility [25]; ordinary EP is the case $\mathbf{S} = \mathbf{I}$

For a parameter in the first oscillator, for example β_1 , the parameter sensitivity is given by

$$\mathbf{F}_{\beta_1}(\mathbf{u}) = \begin{bmatrix} u_1^3 \\ 0 \end{bmatrix}. \quad (17)$$

The corresponding gradient contribution is therefore

$$\frac{d\chi}{d\beta_1} = \int_0^T a_1(T-t) u_1(t)^3 dt. \quad (18)$$

Only the forward and adjoint trajectories of the first DOF are needed. Similarly, $d\chi/dk_1 = \int_0^T a_1(T-t) u_1(t) dt$. The coupling parameter requires both oscillators:

$$\frac{d\chi}{dc} = \int_0^T (a_1(T-t) - a_2(T-t)) (u_1(t) - u_2(t)) dt. \quad (19)$$

These examples illustrate what we mean by a *local learning rule* (e.g. [5, 29, 45]): the gradient with respect to a given parameter depends only on the forward and adjoint dynamics of the degree or degrees of freedom to which this parameter couples. This locality follows because the parameter sensitivity of the internal force $\mathbf{F}_{(\cdot)}$, and likewise of the mass, damping, and drive operators, is supported only on those degrees of freedom. The gradient overlap integral therefore reduces to those degrees of freedom.

Not knowing the parameters. A further simplification arises when the parameters enter the forward operators linearly, as they do in Eq. (16). The sensitivity $\mathbf{F}_{\mathbf{p}_F}$ is then independent of $\mathbf{p}_F$, so the overlap integrals (18) and (19) are assembled entirely from the measured forward and adjoint trajectories and never reference the parameter values themselves. In an experiment, the gradient is thus evaluated at the device's current parameter values, so slow drift, aging, and unknown calibration offsets are absorbed automatically instead of biasing the gradient estimate.

But knowing the model. Knowing which observable to record, and which overlap integral belongs to which parameter, is a statement about the structure of $\mathbf{F}$ and therefore still requires a model, unlike self-learning schemes that avoid explicit knowledge of the detailed energy function or Hamiltonian [5, 26, 70].

5 Second-order real-valued systems

Consider a second-order, real-valued and non-autonomous system introduced in Sec. 2. The explicit time dependence is assumed to be externally programmable, so that it can be replayed as $t \mapsto T - t$ during the adjoint experiment. We wish to shape its trajectory such that it minimizes the cost function (1). The adjoint field required to compute the loss function's gradients in Eqs. (8)–(13) can be obtained on the same system under the following sufficient conditions:

Theorem 5.1 (Sufficient conditions for physical adjoints for second-order real-valued systems). *The adjoint field can be obtained on the same hardware in the following cases:*

- **Linear: Reciprocity.** Suppose

$$\mathbf{F}[\mathbf{u}, \mathbf{p}_F, t] = \mathbf{K}(\mathbf{p}_K, t)\mathbf{u},$$

and reciprocity holds, i.e.,

$$\mathbf{M}^T = \mathbf{M}, \quad \mathbf{D}^T = \mathbf{D}, \quad \mathbf{K}(\mathbf{p}_K, t)^T = \mathbf{K}(\mathbf{p}_K, t), \quad \forall t \in [0, T], \quad (20)$$

then the adjoint is propagated by the same linear hardware, with the explicit time dependence replayed in reverse and with adjoint source (5) and initial conditions (6), (7).

- **Nonlinear: Existence of a Time Reversal Mirror (TRM) and reciprocity of the linearized dynamics.** Suppose the time-reversed trajectory can be generated on the same hardware by an operation on the final state only (applying the TRM), together with the reversal of explicit time dependences. This means initializing with $\mathbf{u}(T)$ and $-\dot{\mathbf{u}}(T)$. If, in addition,

$$\mathbf{D} = \mathbf{0}, \quad \mathbf{M}^T = \mathbf{M}, \quad \mathbf{F}_{\mathbf{u}}[\mathbf{u}(t), \mathbf{p}_F, t]^T = \mathbf{F}_{\mathbf{u}}[\mathbf{u}(t), \mathbf{p}_F, t] \quad \forall t \in [0, T], \quad (21)$$

then the adjoint is obtained as the infinitesimal response of a nudged time-reversed trajectory.

Non-autonomous dynamics and reciprocity. In Theorem 5.1 we allow the internal force $\mathbf{F}$ to be a function of time, i.e. the parameters may now specify a time-dependent schedule rather than being static. The reciprocity condition is instantaneous, i.e. it must hold at every time step, and the explicit time dependence must be replayed in reverse during the adjoint experiment. On the other hand, $\mathbf{M}$ and $\mathbf{D}$ need to be constant in time because otherwise the adjoint state equation will include time derivatives of those matrices and so the forward and adjoint systems will be fundamentally different.

5.1 Linear

The forward dynamics is

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{D}\dot{\mathbf{u}}(t) + \mathbf{K}(t)\mathbf{u}(t) - \mathbf{f}(t) = \mathbf{0}. \quad (22)$$

In the reciprocal case where $\mathbf{M}^T = \mathbf{M}$, $\mathbf{D}^T = \mathbf{D}$, and $\mathbf{K}(t)^T = \mathbf{K}(t)$, the adjoint equation (4) becomes

$$\mathbf{M}\ddot{\mathbf{a}}(t) + \mathbf{D}\dot{\mathbf{a}}(t) + \mathbf{K}(T-t)\mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}. \quad (23)$$

Thus the adjoint experiment has exactly the same *damped* second-order form as the forward experiment. This proves the linear part of Theorem 5.1. $\square$

Physical adjoint algorithm. One reuses the same hardware, sets the applied drive to $-\mathbf{f}_{\text{adj}}$, replaces the initial displacement and velocity by the cost-derived adjoint initial conditions (no nudging required), and replays any explicit time dependence as $t \mapsto T - t$. In this linear reciprocal regime the same damped hardware propagates the adjoint field (see Fig. 2, a).

Remark. If the adjoint source or the initial conditions are too strong, they may excite the system into a large-amplitude regime where the linear equations no longer hold. In that case, the adjoint source and initial conditions can simply be rescaled by a constant, which rescales the norm of the gradient by the same constant.

Literature. In numerical wave-based inverse problems, the same operator governs the forward and adjoint problems when reciprocity holds [28]. In physical learning, reciprocity was used for backpropagation in early work by Hermans et al. [20]. Berneman and Hexner [72] study a contrastive trajectory rule for damped reciprocal systems that uses nudged trajectories and applies to periodic steady states or resting initial conditions. Our linear reciprocal adjoint construction requires no nudging and accommodates arbitrary initial conditions, including gradients with respect to them. Kwon et al. [71] experimentally realize the time-domain adjoint in a coupled RLC network. They write the circuit in first-order voltage-current form; eliminating the currents gives the damped reciprocal second-order dynamics treated here. In Sec. 7 and Sec. 8 we discuss further references that use reciprocity for same-device gradient calculation in Schrödinger (including free-space optics) and stationary (including Helmholtz) problems, respectively.

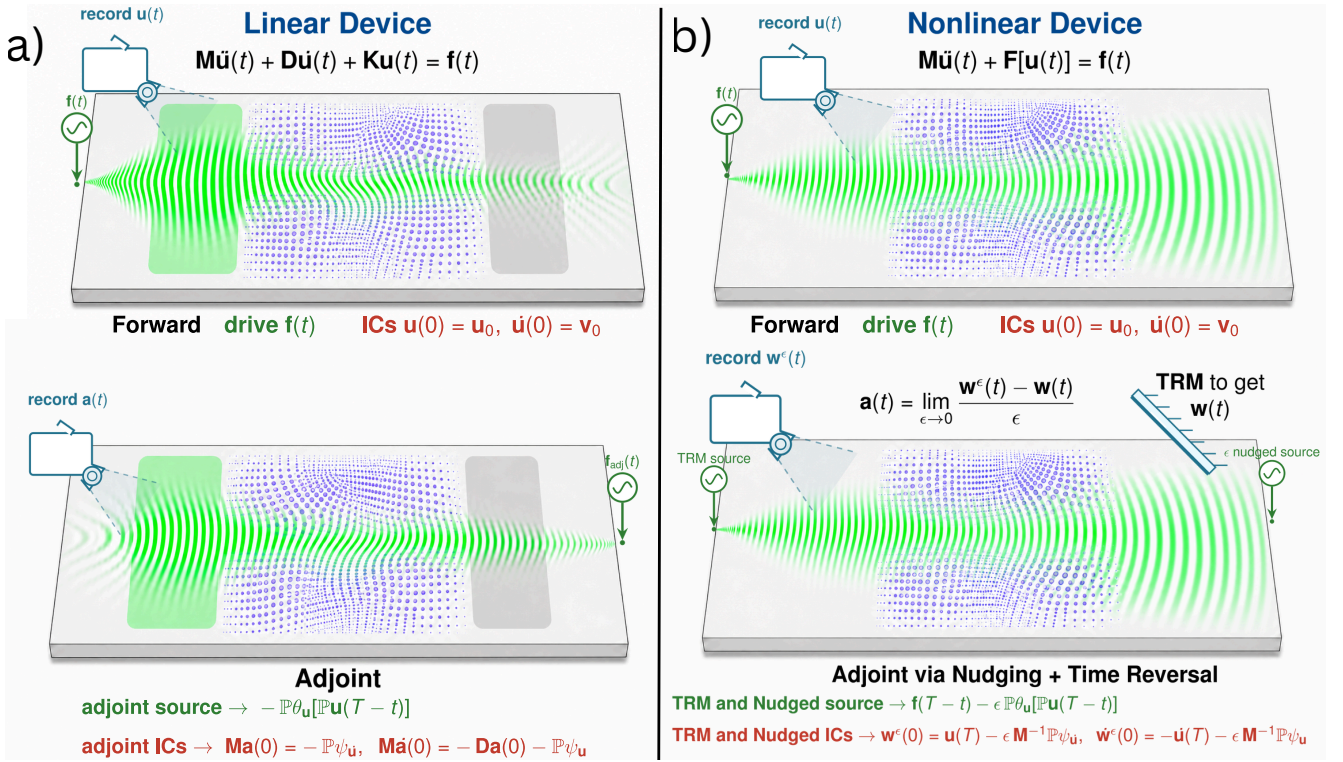


Figure 2: A schematic illustration of the experiments to obtain the adjoint field physically. For the linear devices (a), reciprocity allows the same device to be used, with the drive and ICs adjusted for the forward and adjoint field propagations. Damping and gain are allowed, so long as reciprocity is preserved. In the nonlinear case (b), the existence of a time reversal mirror (TRM) [80] together with reciprocity of the linearized system provides indirect access to the adjoint field in the same hardware. To obtain the adjoint field physically, simultaneously apply the TRM to the final state of the forward run and perturb the time-reversed trajectory with the adjoint drive and ICs. The adjoint field is then obtained as the (infinitesimal) differential response of the reverse trajectory to the nudge.

5.2 Nonlinear

The nonlinear case is more restrictive because the experiment must reproduce the reversed forward trajectory before the infinitesimal nudge can reveal the adjoint. To see the obstruction, set $\mathbf{w}(t) = \mathbf{u}(T-t)$. Evaluating (2) at $T-t$ gives

$$\mathbf{M}\ddot{\mathbf{w}}(t) - \mathbf{D}\dot{\mathbf{w}}(t) + \mathbf{F}[\mathbf{w}(t), \mathbf{p}_F, T-t] - \mathbf{f}(T-t, \mathbf{p}_f) = \mathbf{0}, \quad \mathbf{w}(0) = \mathbf{u}(T), \quad \dot{\mathbf{w}}(0) = -\dot{\mathbf{u}}(T). \quad (24)$$

The damping term has the wrong sign for the same hardware. Reproducing $\mathbf{w}$ would therefore require either $\mathbf{D} = \mathbf{0}$ or an experimental replacement $\mathbf{D} \mapsto -\mathbf{D}$, i.e. turning damping into gain. Even granting such a gain-flipped run does not fix the adjoint problem. We perturb the time-reversed trajectory with the adjoint ICs (6)–(7) and source (5):

$$\mathbf{w}^\epsilon(0) = \mathbf{u}(T) - \epsilon \mathbf{M}^{-T} \mathbb{P}\psi_{\dot{\mathbf{u}}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)], \quad (25)$$

$$\dot{\mathbf{w}}^\epsilon(0) = -\dot{\mathbf{u}}(T) - \epsilon \mathbf{M}^{-T} \left(\mathbf{D}^T \mathbf{a}(0) + \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)] \right) \quad (26)$$

$$\mathbf{f}^\epsilon(t) = \mathbf{f}(T-t, \mathbf{p}_f) - \epsilon \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)], \quad (27)$$

where $|\epsilon| \ll 1$. The perturbation $\delta\mathbf{w} = \mathbf{w}^\epsilon - \mathbf{w}$ around the gain-flipped reversed trajectory due to this small nudge obeys (see Appendix B)

$$\mathbf{M}\delta\ddot{\mathbf{w}}(t) - \mathbf{D}\delta\dot{\mathbf{w}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{w}(t), \mathbf{p}_F, T-t]\delta\mathbf{w}(t) + \epsilon \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] + \mathcal{O}(\epsilon^2) = \mathbf{0}, \quad |\epsilon| \ll 1. \quad (28)$$

The adjoint equation, however, propagates with $+\mathbf{D}^T \dot{\mathbf{a}}$. For reciprocal damping, $\mathbf{D}^T = \mathbf{D}$, the two tangent operators differ by $2\mathbf{D}\dot{\mathbf{a}}$. The same hardware cannot simultaneously reproduce the reversed nonlinear background trajectory and propagate the correct adjoint perturbation around it, except in the undamped case.

For mass-proportional damping, if one can flip damping (gain) to gain (damping), then the adjoint field can be obtained physically and the gradient can be calculated by exponential time weighting with the known damping coefficient [81] (Appendix C). This, however, falls outside our setting, since flipping the sign of the damping changes the hardware between the forward and the adjoint run.

If $\epsilon \rightarrow 0$ and $\mathbf{D} = \mathbf{0}$, the perturbation dynamics (28) recovers the adjoint state equation (4) provided that $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ (reciprocity along the forward trajectory). Then

$$\mathbf{a}(t) = \lim_{\epsilon \rightarrow 0} \frac{\delta \mathbf{w}(t)}{\epsilon} = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(t) - \mathbf{w}(t)}{\epsilon}, \quad (29)$$

which can therefore be extracted from the difference between the nudged and unnudged reversed trajectories (see Eq. (31)). This proves the nonlinear part of Theorem 5.1. $\square$

Remark: only the tangent condition is used. We require $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ only along the visited trajectory; no potential is needed for the construction. If this symmetry holds on an open neighborhood, the Poincaré lemma gives $\mathbf{F} = E_{\mathbf{u}}$ locally there [82]; if it holds throughout a simply connected state space, a single global potential exists. The same distinction applies in the complex case of Sec. 8.2: on an open set, condition (64) is equivalent to closedness of $2 \operatorname{Re}\langle \mathbf{F}, d\mathbf{u} \rangle_{\mathbb{C}}$ and hence locally to (72).

Physical adjoint algorithm. One first runs the forward dynamics and records $\mathbf{u}(t)$; the time-reversed trajectory $\mathbf{w}(t) = \mathbf{u}(T - t)$ is then available directly from this recording and need not be regenerated physically. The adjoint requires only one further experiment, in which the TRM and the nudge are applied *together*: On the same (undamped) hardware, initialize at the final state with the velocity flipped and both initial conditions perturbed in the directions set by the terminal-loss derivatives,

$$\mathbf{w}^\epsilon(0) = \mathbf{u}(T) - \epsilon \mathbf{M}^{-1} \mathbb{P} \psi_{\dot{\mathbf{u}}}, \quad \dot{\mathbf{w}}^\epsilon(0) = -\dot{\mathbf{u}}(T) - \epsilon \mathbf{M}^{-1} \mathbb{P} \psi_{\mathbf{u}}, \quad |\epsilon| \ll 1, \quad (30)$$

drive the system with the reversed external drive superimposed with the adjoint source, $\mathbf{f}(T - t) - \epsilon \mathbb{P} \theta_{\mathbf{u}}[\mathbb{P} \mathbf{u}(T - t)]$, and evolve it forward while replaying the explicit time dependence of the internal forces in reverse, $\mathbf{F}[\cdot, \mathbf{p}_F, t] \mapsto \mathbf{F}[\cdot, \mathbf{p}_F, T - t]$. From the recording of $\mathbf{w}^\epsilon(t)$, the adjoint field is recovered as the infinitesimal differential response

$$\mathbf{a}(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(t) - \mathbf{w}(t)}{\epsilon}, \quad \mathbf{w}(t) = \mathbf{u}(T - t), \quad (31)$$

and the gradient is evaluated digitally from the overlap integrals (8)–(13) (recall $\mathbf{D} = \mathbf{0}$; see Fig. 2, b).

Damping compensation with external forces. The obstruction above can be bypassed if the damping matrix or gain matrix $\mathbf{D}$ is known and one can apply additional forces at each damped DOF during the nudged evolution. To do so, one applies an additional drive computed from $\mathbf{D}$ and the recorded, time-reversed velocity:

$$\mathbf{f}^\epsilon(t) = \mathbf{f}(T - t, \mathbf{p}_f) - \epsilon \mathbb{P} \theta_{\mathbf{u}}[\mathbb{P} \mathbf{u}(T - t)] + 2\mathbf{D}\dot{\mathbf{w}}(t). \quad (32)$$

Then the perturbation evolves according to

$$\mathbf{M} \delta \ddot{\mathbf{w}} + \mathbf{D} \delta \dot{\mathbf{w}} + \mathbf{F}_{\mathbf{u}}[\mathbf{w}] \delta \mathbf{w} + \mathbb{P} \theta_{\mathbf{u}} + \mathcal{O}(\epsilon^2) = \mathbf{0}. \quad (33)$$

Thus, if $\mathbf{M}^T = \mathbf{M}$, $\mathbf{D}^T = \mathbf{D}$, and $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ along the trajectory, the differential response recovers the forward-time adjoint even in the presence of damping. The construction therefore remains a same-device adjoint in our sense because it leaves the intrinsic dynamics unchanged. The compensating force

$2\mathbf{D}\dot{\mathbf{w}}(t)$ must be computed from the recorded trajectory and applied to every degree of freedom on which $\mathbf{D}$ acts. This requires accurate knowledge of $\mathbf{D}$, distributed actuation, and precise trajectory replay. Moreover, the compensation must be prescribed from the reference trajectory; applying it as feedback to the auxiliary trajectory would modify the tangent dynamics. The same damping compensation was used in the concurrent work by Li et al. [79] for the nonlinear Schrödinger equation in combination with $\mathcal{PT}$ -symmetry (see Sec. 9.2). We exclude such compensation from Table 1, because it requires calibrated external force control at every damped DOF. It may nevertheless be practical when only a few DOFs are damped and their damping is known.

Symmetric nudging. At finite ϵ the extraction (31) is a forward difference, so the measured field is the adjoint up to an $\mathcal{O}(\epsilon)$ bias from the second-order term of the perturbation expansion. Running instead two nudged experiments with opposite signs and forming the central difference,

$$\mathbf{a}(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^{+\epsilon}(t) - \mathbf{w}^{-\epsilon}(t)}{2\epsilon}, \quad (34)$$

leaves an $\mathcal{O}(\epsilon^2)$ bias, at the cost of one additional experiment [34, 83].

Literature. For time-reversible Hamiltonian dynamics, the construction above implements the echo principle of Hamiltonian echo backpropagation [26] and its recurrent extension by Pourcel and Ernault [34]. The latter was shown to be equivalent to the continuous adjoint method; the self-adaptation of the physical parameters proposed by López-Pastor and Marquardt [26], in which the parameters themselves become dynamical, is outside our scope. Our formulation also covers time-dependent parameters of the internal force. A complementary, variational route recasts the learning signal through the classical action: Maszar [84] and Pourcel et al. [81] generalize Equilibrium Propagation [29] from energy-based fixed points to full trajectories (Lagrangian Equilibrium Propagation), replacing its two-state energy contrast with a contrast of Lagrangian parameter-derivatives integrated over the trajectory. Because this contrast follows from an integration by parts, it carries boundary terms. Depending on the imposed boundary conditions, these terms either vanish by construction (e.g. fixed or periodic endpoints) or persist as generally intractable residuals. For time-reversible systems, Pourcel et al. [81] identify a Parametric Final Value Problem boundary choice that renders these residuals tractable and recovers the recurrent Hamiltonian echo rule [34] as the physically practical instantiation of this variational construction.

5.3 Summary: linear and nonlinear systems behave fundamentally differently

We have shown that linear systems require weaker conditions and allow simpler finite-amplitude physical adjoint field constructions (Fig. 2). The simplicity of physical adjoint experiments for linear systems makes a strong case for parameter encoding (structural nonlinearities) [53, 59, 60] and source-position encoding [58], which give the linear system nonlinear computational capabilities. Yet physical gradients are readily accessible with finite-amplitude constructions and in the realistic damped setting [36]. Damping and gain parameters, whose gradients are available once the adjoint field is obtained (Eq. (9)), can also be used to increase the expressivity of the physical device.

6 First-order real-valued systems

We next consider first-order, real-valued¹ and non-autonomous dynamics

$$\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}. \quad (35)$$

¹Here, we consider overdamped first-order systems. First-order reformulations of second-order systems that retain both position and velocity are covered in Sec. 5.

The corresponding adjoint equation, derived in Appendix E, is

$$\dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\mathbf{a}(t) + \mathbf{f}_{\text{adj}}(t) = \mathbf{0}, \quad (36)$$

with adjoint source and initial conditions given by

$$\mathbf{f}_{\text{adj}}(t) = \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)], \quad \mathbf{a}(0) = -\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T)]. \quad (37)$$

Again, we assume external control of the explicit time dependences of $\mathbf{F}$ and $\mathbf{f}$, so that they can be replayed as $t \mapsto T-t$ during the adjoint experiment.

Theorem 6.1 (Sufficient conditions for physical adjoints for first-order real-valued systems). *The adjoint field can be obtained on the same hardware in the following cases:*

- **Linear: Reciprocity.** Suppose

$$\mathbf{F}[\mathbf{u}, \mathbf{p}_F, t] = \mathbf{K}(\mathbf{p}_K, t)\mathbf{u}$$

and reciprocity holds, i.e.,

$$\mathbf{K}(\mathbf{p}_K, t)^T = \mathbf{K}(\mathbf{p}_K, t), \quad \forall t \in [0, T], \quad (38)$$

then the adjoint is propagated by the same linear hardware, with the explicit time dependence replayed in reverse and with the adjoint source and initial conditions (37).

- **Nonlinear, overdamped: trajectories are obstructed without prescribed compensation, the steady state is not.**

6.1 Linear

The forward dynamics is

$$\dot{\mathbf{u}}(t) + \mathbf{K}(t)\mathbf{u}(t) - \mathbf{f}(t) = \mathbf{0}. \quad (39)$$

If the operator is reciprocal, $\mathbf{K}(t)^T = \mathbf{K}(t)$, the adjoint equation (36) becomes

$$\dot{\mathbf{a}}(t) + \mathbf{K}(T-t)\mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}. \quad (40)$$

Thus the adjoint experiment has exactly the same overdamped first-order form as the forward experiment and can be performed on the same hardware. $\square$

Physical adjoint algorithm. One reuses the same hardware, sets the applied drive to $-\mathbf{f}_{\text{adj}}$, replaces the initial condition by the loss-derived adjoint initial condition, and replays any explicit time dependence as $t \mapsto T-t$.

Literature. Hermans et al. [20] formulated physical backpropagation using reciprocal impulse responses, explicitly including linear heat and diffusion equations. Here we recover this construction in the first-order adjoint formulation and extend it to explicitly time-dependent coefficients. Berneman and Hexner [73] proposed a Rayleighian learning rule that is approximate for general nonlinear trajectories.

6.2 Nonlinear

The first-order ODE of this section is the large-damping limit of the second-order ODE (Sec. 5). The nonlinear adjoint construction is therefore obstructed for the same reason discussed there: time reversal requires flipping damping into gain (here $\mathbf{F} \rightarrow -\mathbf{F}$) while the adjoint requires damping. Trajectory gradients of nonlinear overdamped dynamics are therefore not accessible through this reversed-drive construction.

Trajectory compensation with external forces. For $\mathbf{w}(t) = \mathbf{u}(T - t)$, the prescribed drive $\mathbf{f}^\epsilon(t) = \mathbf{f}(T - t) + 2\dot{\mathbf{w}}(t) - \epsilon \mathbf{f}_{\text{adj}}(t)$ and initial state $\mathbf{w}^\epsilon(0) = \mathbf{u}(T) + \epsilon \mathbf{a}(0)$ yield the adjoint as an infinitesimal response when $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ along the trajectory. This requires accurate velocity recording and external force control at each DOF with nonzero reference velocity, and is excluded from Table 1.

Without trajectory compensation, the steady state remains accessible. If $\mathbf{F}$ and $\mathbf{f}$ are time-independent and the dynamics converges to a linearly exponentially stable fixed point, the loss depends on the parameters only through that fixed point and the problem becomes stationary. That case is Equilibrium Propagation, treated in Sec. 8.2; its relation to the trajectory adjoint of this section is derived in Appendix I.3.

7 Schrödinger-type complex systems

We now consider complex fields governed by

$$i\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, \quad \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}). \quad (41)$$

Throughout this section derivatives are Wirtinger derivatives [85, 86]: for $z = x + iy$, $\partial_z = \frac{1}{2}(\partial_x - i\partial_y)$ and $\partial_{\bar{z}} = \frac{1}{2}(\partial_x + i\partial_y)$, and for vector fields we treat $\mathbf{u}$ and its complex conjugate $\bar{\mathbf{u}}$ as formally independent variables. Thus $\mathbf{F}_{\mathbf{u}}$ and $\mathbf{F}_{\bar{\mathbf{u}}}$ denote the two Wirtinger Jacobians, and for the real-valued cost functions the corresponding derivatives satisfy $\theta_{\mathbf{u}} = \bar{\theta}_{\bar{\mathbf{u}}}$ and $\psi_{\mathbf{u}} = \bar{\psi}_{\bar{\mathbf{u}}}$ and similarly for χ . As we will see, the physical adjoint construction delivers the complex conjugate of the adjoint field. Writing $\mathbf{c}(t) = \overline{\mathbf{b}(T - t)}$, the forward-time complex conjugate adjoint equation derived in Appendix F is

$$i\dot{\mathbf{c}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T - t), \bar{\mathbf{u}}(T - t), \mathbf{p}_F, T - t]\mathbf{c}(t) + \overline{\mathbf{F}_{\bar{\mathbf{u}}}[\mathbf{u}(T - t), \bar{\mathbf{u}}(T - t), \mathbf{p}_F, T - t]}^T \bar{\mathbf{c}}(t) + \mathbf{f}_{\text{adj}}(t) = \mathbf{0}, \quad (42)$$

where $\overline{(\cdot)}^T$ denotes the conjugate transpose. The source and initial conditions are

$$\mathbf{f}_{\text{adj}}(t) = \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T - t), \mathbb{P}\bar{\mathbf{u}}(T - t)], \quad \mathbf{c}(0) = i \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\bar{\mathbf{u}}(T)]. \quad (43)$$

Taking the complex conjugate of Eq. (42) flips the signs in front of the Wirtinger Jacobians, which is why $\mathbf{c}$, rather than $\mathbf{b}$, can be obtained physically in the same hardware.

Inner products now include complex conjugation, $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{C}} = \bar{\mathbf{x}}^T \mathbf{y}$. The total variation of χ is therefore $\delta\chi = \langle d\chi/d\bar{\mathbf{u}}, \delta\mathbf{u} \rangle_{\mathbb{C}} + \langle d\chi/d\mathbf{u}, \delta\bar{\mathbf{u}} \rangle_{\mathbb{C}} = 2 \operatorname{Re} \langle d\chi/d\bar{\mathbf{u}}, \delta\mathbf{u} \rangle_{\mathbb{C}}$. The gradient overlap integrals expressed in terms of $\mathbf{c}$ then become (Appendix F):

$$\frac{d\chi}{d\mathbf{p}_F} = 2 \operatorname{Re} \int_0^T \left\langle \overline{\mathbf{c}(T - t)}, \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t] \right\rangle_{\mathbb{C}} dt, \quad (44)$$

$$\frac{d\chi}{d\mathbf{p}_f} = -2 \operatorname{Re} \int_0^T \left\langle \overline{\mathbf{c}(T - t)}, \mathbf{f}_{\mathbf{p}_f}(t) \right\rangle_{\mathbb{C}} dt, \quad (45)$$

$$\frac{d\chi}{d\mathbf{p}_{u_0}} = -2 \operatorname{Re} \left\langle \overline{i\mathbf{c}(T)}, \mathbf{u}_0, \mathbf{p}_{u_0} \right\rangle_{\mathbb{C}}. \quad (46)$$

The sufficient conditions under which $\mathbf{c}$ can be obtained in the same hardware are as follows:

Theorem 7.1 (Sufficient conditions for physical adjoints for Schrödinger-type systems). *The complex conjugate of the adjoint field can be obtained on the same hardware in the following cases:*

- **Linear: Reciprocity.** Suppose

$$\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t] = \mathbf{K}(\mathbf{p}_K, t)\mathbf{u}$$

and reciprocity holds, i.e.,

$$\mathbf{K}(\mathbf{p}_K, t)^T = \mathbf{K}(\mathbf{p}_K, t), \quad \forall t \in [0, T],$$

then the complex conjugate of the adjoint is propagated by the same linear hardware, with the explicit time dependence replayed in reverse, the applied drive $-\mathbf{f}_{\text{adj}}$, and the initial condition in Eq. (43).

- **Nonlinear: Conjugation TRM and reciprocity of the linearized dynamics.** Let $\mathcal{K}$ denote complex conjugation. Suppose $\mathbf{F}$ is continuously Wirtinger differentiable on a conjugation-invariant operating domain and obeys

$$\mathbf{F}[\bar{\mathbf{u}}, \mathbf{u}, \mathbf{p}_F, t] = \overline{\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t]}, \quad \forall t \in [0, T]. \quad (47)$$

Then $\mathcal{K}$ is a TRM: the reversed background $\mathbf{w}(t) = \overline{\mathbf{u}(T-t)}$ is generated on the same hardware by conjugating the final state, reversing the explicit time dependences, and applying the drive $\bar{\mathbf{f}}(T-t)$. Suppose, independently, that along the forward trajectory

$$\mathbf{F}_{\mathbf{u}}^T = \overline{\mathbf{F}_{\mathbf{u}}}, \quad (48)$$

$$\overline{\mathbf{F}_{\mathbf{u}}}^T = \mathbf{F}_{\bar{\mathbf{u}}}, \quad \forall t \in [0, T], \quad (49)$$

where both Jacobians are evaluated at $[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t]$. Then $\mathbf{c}$ is the infinitesimal response of a nudged conjugation-reversed trajectory.

7.1 Linear

The linear case requires symmetry in the reciprocal sense, not Hermiticity: loss and gain are compatible with the physical adjoint as long as reciprocity is preserved. For a complex-linear system, the forward dynamics is

$$i\dot{\mathbf{u}}(t) + \mathbf{K}(\mathbf{p}_K, t)\mathbf{u}(t) - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, \quad (50)$$

where $\mathbf{F}_{\bar{\mathbf{u}}} = \mathbf{0}$ by complex linearity and $\mathbf{F}_{\mathbf{u}} = \mathbf{K}(\mathbf{p}_K, t)$ is independent of the forward state. Then Eq. (42) reads

$$i\dot{\mathbf{c}}(t) + \mathbf{K}(\mathbf{p}_K, T-t)^T \mathbf{c}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t), \mathbb{P}\bar{\mathbf{u}}(T-t)] = \mathbf{0}. \quad (51)$$

Hence, if $\mathbf{K}^T = \mathbf{K}$ for all t , one can propagate both fields in the same device by setting the applied drive to $-\mathbf{f}_{\text{adj}}$, imposing the adjoint initial condition, and replaying the explicit time dependence in reverse. $\square$

Free-space diffractive optical systems. In Appendix H we show that replacing time t by the propagation coordinate z recovers free-space diffractive optical systems.

Holographic vs. intensity measurements. The physical algorithm above requires measuring both $\mathbf{u}(t)$ and $\mathbf{c}(t)$ holographically (phase and amplitude), so that the gradient overlap integrals (44)–(46) can be evaluated. Extracting the gradient from intensity measurements alone would be experimentally simpler. In [23, 35] this was achieved for Helmholtz systems by exploiting time reversal symmetry in addition to reciprocity. The same construction applies to linear Schrödinger problems:

For two complex numbers x and y , $2\text{Re}(\bar{x}y) = |x+y|^2 - |x|^2 - |y|^2$, so $\text{Re}(\bar{x}y)$ follows from three intensity measurements. Eqs. (44)–(46) are instead of the form $\text{Re}(xy)$, with x not conjugated, which intensities do not give; one needs $\mathbf{b}$ physically in the system rather than $\mathbf{c}(t) = \overline{\mathbf{b}(T-t)}$. This can be achieved if the system is time reversal symmetric and reciprocal: (i) evolve the forward state $\mathbf{u}$ from 0 to T and measure the intensity $I_{\mathbf{u}}(t) = |\mathbf{u}(t)|^2$ and (ii) physically evolve $\mathbf{c}$ from 0 to T and measure the intensity $I_{\mathbf{c}}(t) = |\mathbf{c}(t)|^2$. Then, (iii), apply the TRM of complex conjugating $\mathbf{c}(T) \rightarrow \bar{\mathbf{c}}(T)$, time reverse the adjoint

source and add the forward sources and initial conditions so that the system now evolves the forward state plus the actual adjoint state $\mathbf{u}(t) + \overline{\mathbf{c}(T-t)} = \mathbf{u}(t) + \mathbf{b}(t)$; and record $I_{\mathbf{u}+\mathbf{b}}(t) = |\mathbf{u}(t) + \mathbf{b}(t)|^2$. The instantaneous overlap can then be constructed as $I_{\mathbf{u}+\mathbf{b}}(t) - I_{\mathbf{u}}(t) - I_{\mathbf{c}}(T-t)$. The construction requires time reversal symmetry, so that complex conjugating $\mathbf{c}(T)$ regenerates the adjoint state evolution, and linearity, since step (iii) uses the superposition principle.

This gives the site-wise overlap $2 \operatorname{Re}[\overline{b_i(t)} u_i(t)]$, and time reversal symmetry makes $\mathbf{K}$, and with it $\mathbf{K}_{\mathbf{p}}$, real. Eq. (44) therefore follows when $\mathbf{K}_{\mathbf{p}}$ is diagonal in the measurement basis, as for an on-site potential, but not for a coupling coefficient, which needs the cross-site overlaps $2 \operatorname{Re}[\overline{b_i} u_j]$. The source and initial-condition gradients (45)–(46) pair $\mathbf{b}$ with a known vector rather than with $\mathbf{u}$ and still require holographic measurements.

Literature. For phase-only spatial light modulators (SLMs), the framework reduces to the in situ optical backpropagation rule theorized by Zhou et al. [21]. Wagner and Psaltis proposed the same layerwise adjoint structure in a holographic free-space optical network: reciprocal volume holograms propagated the transpose of each linear interconnection, while a weak counterpropagating probe through nonlinear Fabry-Pérot neurons was designed to approximate multiplication by the local activation derivative [39]. This is a direct small-signal implementation of the nonlinear tangent rather than a contrastive nudging construction; because the probe transmission only approximates the required derivative, the resulting physical gradient is approximate through the nonlinear units. The same reciprocal linear-adjoint construction covers the linear optical layers proposed in Guo et al. [40] and experimentally realized in Spall et al. [41]; in both cases, the saturable-absorber response only approximates the nonlinear tangent required by backpropagation, so the construction is exact for the ideal linear layers but approximate through the nonlinearity. Mididoddi et al. [22] obtained the adjoint field experimentally to optimize the input field and suppress output fluctuations in a dynamically varying medium. In the linear case, the adjoint source can be applied at full amplitude and no nudging or difference of trajectories is required.

7.2 Nonlinear

Nonlinear Schrödinger systems require two independent ingredients. A TRM must regenerate the reversed background, and the tangent dynamics about that background must match the adjoint dynamics. Here the TRM is ordinary spinless time reversal, implemented by complex conjugation.

Let

$$\mathbf{w}(t) := \overline{\mathbf{u}(T-t)}. \quad (52)$$

Under condition (47), this field obeys

$$\begin{aligned} i\dot{\mathbf{w}}(t) + \mathbf{F}[\mathbf{w}(t), \overline{\mathbf{w}(t)}, \mathbf{p}_F, T-t] - \overline{\mathbf{f}(T-t, \mathbf{p}_f)} &= \mathbf{0}, \\ \mathbf{w}(0) &= \overline{\mathbf{u}(T)}. \end{aligned} \quad (53)$$

Thus conjugating the final state and the drive, while reversing every explicit time dependence, regenerates the reversed background on the same hardware.

Let

$$\mathbf{q}(t) := \mathbf{f}_{\text{adj}}(t), \quad \mathbf{q}_T := \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\overline{\mathbf{u}(T)}].$$

Run a nudged copy with

$$\begin{aligned} i\dot{\mathbf{w}}^\epsilon(t) + \mathbf{F}[\mathbf{w}^\epsilon(t), \overline{\mathbf{w}^\epsilon(t)}, \mathbf{p}_F, T-t] - [\overline{\mathbf{f}(T-t, \mathbf{p}_f)} - \epsilon\mathbf{q}(t)] &= \mathbf{0}, \\ \mathbf{w}^\epsilon(0) &= \mathbf{w}(0) + \epsilon i\mathbf{q}_T. \end{aligned} \quad (54)$$

The physical nudge is $-\epsilon\mathbf{q}$ because the external drive enters the residual with a minus sign. Differentiating condition (47) gives the tangent operators about $\mathbf{w}$ in terms of those along the forward trajectory.

Conditions (48) and (49) then make the differential response obey Eq. (42). Therefore

$$\mathbf{c}(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(t) - \mathbf{w}(t)}{\epsilon}, \quad \mathbf{b}(t) = \overline{\mathbf{c}(T - t)}. \quad (55)$$

The construction is exact in the infinitesimal-nudge limit. For a sufficiently smooth response, a one-sided difference has an $\mathcal{O}(\epsilon)$ bias, while two opposite nudges reduce it to $\mathcal{O}(\epsilon^2)$. Appendix F.4 gives the calculation in the more general setting used in Sec. 9. $\square$

Example: local Kerr nonlinearity. Consider

$$\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}] = \mathbf{K}\mathbf{u} + g|\mathbf{u}|^2 \odot \mathbf{u}, \quad g \in \mathbb{R}. \quad (56)$$

Conjugation is a TRM when $\mathbf{K} = \bar{\mathbf{K}}$. The Wirtinger Jacobians are

$$\mathbf{F}_{\mathbf{u}} = \mathbf{K} + 2g \operatorname{diag}(|\mathbf{u}|^2), \quad \mathbf{F}_{\bar{\mathbf{u}}} = g \operatorname{diag}(\mathbf{u}^2). \quad (57)$$

The local terms satisfy the tangent conditions automatically. The linear term also requires $\mathbf{K}^T = \mathbf{K}$. Thus ordinary conjugation time reversal gives the physical adjoint for a real symmetric $\mathbf{K}$ and a real local Kerr coefficient. Generalized mirrors and tangent intertwiners are treated in Sec. 9.

Literature. For autonomous Hamiltonian systems with a terminal-value cost, the nudged time-reversed echo dynamics of López-Pastor and Marquardt [26] transports the output perturbation backward and yields the continuous-adjoint gradient. That work also constructs a self-adapting system in which the trainable parameters are physical degrees of freedom updated by the same dynamics, without external gradient computation.

8 Stationary problems and (quantum) Equilibrium Propagation

In the preceding sections, the computation was carried out along a trajectory. We now turn to stationary problems, in which the result is read from a steady state, such as a fixed point of relaxational dynamics or the steady state of a harmonically driven system. We allow complex amplitudes $\mathbf{u} \in \mathbb{C}^N$. We consider the stationary problem

$$\mathbf{F}[\mathbf{u}^*, \bar{\mathbf{u}}^*, \mathbf{p}_F] - \mathbf{f}(\mathbf{p}_f) = \mathbf{0}, \quad (58)$$

where $\mathbf{u}^*$ denotes the steady state, and the cost function is defined on it as

$$\chi = \psi[\mathbb{P}\mathbf{u}^*, \mathbb{P}\bar{\mathbf{u}}^*]. \quad (59)$$

The adjoint field $\mathbf{a} \in \mathbb{C}^N$ associated with Eq. (58) satisfies the following equation (see Appendix I), with both Wirtinger Jacobians evaluated at the steady state:

$$\overline{\mathbf{F}_{\mathbf{u}}}^T \mathbf{a} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{a}} + \mathbf{f}_{\text{adj}} = \mathbf{0}, \quad (60)$$

where the adjoint source is given by

$$\mathbf{f}_{\text{adj}} = \mathbb{P} \psi_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}^*, \mathbb{P}\bar{\mathbf{u}}^*]. \quad (61)$$

Once $\mathbf{a}$ is known, the gradients are a single overlap of the two static fields,

$$\frac{d\chi}{d\mathbf{p}_F} = 2 \operatorname{Re} \langle \mathbf{a}, \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}^*, \bar{\mathbf{u}}^*, \mathbf{p}_F] \rangle_{\mathbb{C}}, \quad (62)$$

$$\frac{d\chi}{d\mathbf{p}_f} = -2 \operatorname{Re} \langle \mathbf{a}, \mathbf{f}_{\mathbf{p}_f}(\mathbf{p}_f) \rangle_{\mathbb{C}}. \quad (63)$$

The sufficient conditions mirror those of the trajectory case. The nonlinear construction nudges the steady state instead of a time-reversed trajectory:

Theorem 8.1 (Sufficient conditions for physical adjoints in stationary problems). *The adjoint field, or its complex conjugate, can be obtained on the same hardware in the following cases:*

- **Linear: Reciprocity.** Suppose

$$\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F] = \mathbf{K}(\mathbf{p}_K) \mathbf{u},$$

and reciprocity holds, i.e.,

$$\mathbf{K}(\mathbf{p}_K)^T = \mathbf{K}(\mathbf{p}_K),$$

then the complex conjugate of the adjoint field, $\mathbf{c} = \bar{\mathbf{a}}$, is generated by solving the same stationary problem with the forward source $\mathbf{f}$ replaced by the adjoint source $-\mathbb{P}\psi_{\mathbf{u}}$, in a single finite-amplitude experiment and with no nudging.

- **Nonlinear: Reciprocity of the tangent operators at the steady state.** Suppose the steady state $\mathbf{u}^*$ is a linearly exponentially stable fixed point of a relaxation realized by the same hardware, and that the full real-linear tangent operator of the stationary equation at $\mathbf{u}^*$ is invertible. If the Wirtinger tangent operators there satisfy

$$\overline{\mathbf{F}_{\mathbf{u}}[\mathbf{u}^*, \mathbf{p}_F]}^T = \mathbf{F}_{\mathbf{u}}[\mathbf{u}^*, \mathbf{p}_F], \quad \mathbf{F}_{\bar{\mathbf{u}}}[\mathbf{u}^*, \mathbf{p}_F]^T = \mathbf{F}_{\bar{\mathbf{u}}}[\mathbf{u}^*, \mathbf{p}_F], \quad (64)$$

then the adjoint field $\mathbf{a}$ is the infinitesimal response of the fixed point under an adjoint source nudge. If instead

$$\mathbf{F}_{\mathbf{u}}[\mathbf{u}^*, \mathbf{p}_F]^T = \mathbf{F}_{\mathbf{u}}[\mathbf{u}^*, \mathbf{p}_F], \quad \overline{\mathbf{F}_{\bar{\mathbf{u}}}[\mathbf{u}^*, \mathbf{p}_F]}^T = \mathbf{F}_{\bar{\mathbf{u}}}[\mathbf{u}^*, \mathbf{p}_F], \quad (65)$$

then $\mathbf{c} = \bar{\mathbf{a}}$ can be obtained on the same device using the complex conjugate adjoint-source nudge.

8.1 Linear

For a linear stationary problem

$$\mathbf{K}(\mathbf{p}_K) \mathbf{u}^* = \mathbf{f}(\mathbf{p}_f), \quad (66)$$

with $\mathbf{K}$ invertible, the internal force $\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F] = \mathbf{K}(\mathbf{p}_K) \mathbf{u}$ is complex-linear, so its Wirtinger Jacobians are constant, $\mathbf{F}_{\mathbf{u}} = \mathbf{K}(\mathbf{p}_K)$ and $\mathbf{F}_{\bar{\mathbf{u}}} = \mathbf{0}$. An example is a harmonically driven system at frequency ω , for which $\mathbf{K} = -\omega^2 \mathbf{M} + i\omega \mathbf{D} + \mathbf{K}_{\text{stat}}$. The adjoint problem then becomes the Hermitian-adjoint system

$$\overline{\mathbf{K}(\mathbf{p}_K)}^T \mathbf{a} = -\mathbb{P} \psi_{\bar{\mathbf{u}}}[\mathbb{P} \mathbf{u}^*, \mathbb{P} \bar{\mathbf{u}}^*], \quad (67)$$

whose conjugate reads

$$\mathbf{K}(\mathbf{p}_K)^T \mathbf{c} = -\mathbb{P} \psi_{\mathbf{u}}[\mathbb{P} \mathbf{u}^*, \mathbb{P} \bar{\mathbf{u}}^*], \quad \mathbf{c} = \bar{\mathbf{a}}. \quad (68)$$

This is the same system if $\mathbf{K}(\mathbf{p}_K)^T = \mathbf{K}(\mathbf{p}_K)$. Hence, the adjoint field can be obtained by swapping the external drive for the adjoint source. $\square$

Comparison with EP. This is the linear, stationary (and complex-valued) counterpart of Equilibrium Propagation (Sec. 8.2). Because the system is linear, the adjoint source is applied at finite amplitude in a single experiment, without nudging. Damping or gain are permissible: reciprocity is required, but Hermiticity is not.

Holographic vs. intensity measurements. For complex amplitudes, holographic measurements are required. Two distinct cases allow the gradients to be reconstructed from intensity measurements.

(i) If $\mathbf{K}$ is real-valued (and hence reciprocity implies Hermiticity), injecting $-\mathbb{P}\psi_{\bar{\mathbf{u}}}$ in place of $-\mathbb{P}\psi_{\mathbf{u}}$ propagates $\mathbf{a}$ itself rather than $\mathbf{c} = \bar{\mathbf{a}}$. Three intensity distributions, $I_{\mathbf{u}^*}$, $I_{\mathbf{a}}$ and $I_{\mathbf{u}^*+\mathbf{a}}$, the last with the applied drive $\mathbf{f} - \mathbb{P}\psi_{\bar{\mathbf{u}}}$, then deliver the site-wise overlaps $2\text{Re}[\bar{a}_i u_i^*] = I_{\mathbf{u}^*+\mathbf{a},i} - I_{\mathbf{u}^*,i} - I_{\mathbf{a},i}$. These evaluate (62) whenever $\mathbf{K}_{\mathbf{p}}$ is real and diagonal in the measurement basis, as for a local permittivity; a non-local $\mathbf{K}_{\mathbf{p}}$ also needs the cross-site overlaps $2\text{Re}[\bar{a}_i u_j^*]$, which intensity distributions alone do not provide.

(ii) If $\mathbf{K}$ is complex-valued and propagation is reciprocal and lossless throughout the device, the analogous procedure of Sec. 7, paragraph “Holographic vs. intensity measurements,” applies. It was developed by Hughes et al. [35] and realized by Pai et al. [23] in silicon photonic meshes of Mach-Zehnder interferometers and programmable phase shifters. The conjugated adjoint field $\mathbf{c}$ is propagated from the outputs to the inputs, measured holographically, and re-encoded as complex-conjugated input sources, so that a further forward pass carries $\mathbf{a}$ rather than $\mathbf{c}$; superimposing it on the forward field gives $I_{\mathbf{u}^*+\mathbf{a}}$. Here $\mathbf{K}_{\mathbf{p}}$ may be complex on the diagonal, as for a phase shifter, which also requires $\text{Im}[\bar{a}_i u_i^*]$. Re-encoding with an extra factor i carries $i\mathbf{a}$, for which $\text{Re}[i\bar{a}_i u_i^*] = \text{Im}[\bar{a}_i u_i^*]$ and $I_{i\mathbf{a}} = I_{\mathbf{a}}$, so a single further distribution $I_{\mathbf{u}^*+i\mathbf{a}}$ suffices.

Literature. Physical backpropagation in Helmholtz systems was theorized for feed-forward photonic neural networks by Hughes et al. [35] and experimentally demonstrated by Pai et al. [23]. In these optical–electronic–optical (OEO) architectures, the optical output of each layer is detected, transformed by the electronic nonlinearity, and re-encoded for the next layer. During backpropagation, the field returned by each optical layer is multiplied electronically by the derivative of the intervening nonlinearity and re-encoded as the adjoint input to the preceding layer, allowing the single-layer protocol to be chained across layers (see Appendix I.4). Ashtiani et al. [87] demonstrated a different on-chip backpropagation architecture with separate forward and backward blocks and approximate physical nonlinear derivatives. For general wave-scattering systems, the in situ physical adjoint computing protocol of Guillamon et al. [37] is equivalent to the holographic formulation presented here; a mechanical equivalent was developed by Li and Mao [24, 38]. Closely related identities were derived by Wanjura and Marquardt [36] for neuromorphic systems based on linear wave scattering. As shown in Appendix I.5, their scattering-matrix gradient formulas follow from the present adjoint identity.

8.2 Nonlinear: Equilibrium Propagation

Let $\mathbf{u}^{*\epsilon}$ be the steady state reached under the drive $\mathbf{f} - \epsilon\mathbb{P}\psi_{\bar{\mathbf{u}}}$,

$$\mathbf{F}[\mathbf{u}^{*\epsilon}, \bar{\mathbf{u}}^{*\epsilon}, \mathbf{p}_F] - \mathbf{f}(\mathbf{p}_F) + \epsilon\mathbb{P}\psi_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}^{*\epsilon}, \mathbb{P}\bar{\mathbf{u}}^{*\epsilon}] = \mathbf{0}, \quad |\epsilon| \ll 1. \quad (69)$$

Subtracting the free steady-state equation and linearizing in $\delta\mathbf{u} = \mathbf{u}^{*\epsilon} - \mathbf{u}^*$ gives

$$\mathbf{F}_{\mathbf{u}}\delta\mathbf{u} + \mathbf{F}_{\bar{\mathbf{u}}}\bar{\delta\mathbf{u}} + \epsilon\mathbb{P}\psi_{\bar{\mathbf{u}}} = \mathcal{O}(\epsilon^2). \quad (70)$$

Divide by ϵ and let $\epsilon \rightarrow 0$. The rescaled response then obeys the adjoint equation (60) precisely when the tangent operators satisfy condition (64), and by uniqueness

$$\mathbf{a} = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{u}^{*\epsilon} - \mathbf{u}^*}{\epsilon}. \quad (71)$$

This proves that conditions (64) are sufficient to obtain $\mathbf{a}$. Taking the complex conjugate of the adjoint equation (60) shows that conditions (65) are sufficient to obtain $\mathbf{c}$. This proves the nonlinear part of Theorem 8.1. $\square$

Physical adjoint algorithm. The device is run twice and the two steady states are subtracted. (i) *Free phase*: let the system settle under the forward drive alone, record $\mathbf{u}^*$, and evaluate the loss derivative from the measured output. (ii) *Nudged phase*: let it settle again under the drive $\mathbf{f} - \epsilon \mathbb{P}\psi_{\bar{\mathbf{u}}}$, and record $\mathbf{u}^{*\epsilon}$ to extract the adjoint field via Eq. (71). The free and nudged phases must settle onto the same branch: ϵ must be small enough that the nudge neither crosses a basin boundary nor drives the system through a bifurcation.

Complex-valued example. Condition (64) holds identically whenever the force is the complex gradient of a real energy,

$$\mathbf{F} = E_{\bar{\mathbf{u}}}, \quad E \in \mathbb{R}, \quad (72)$$

the complex counterpart of $\mathbf{F} = E_{\mathbf{u}}$: reality of E makes $\mathbf{F}_{\mathbf{u}} = E_{\bar{\mathbf{u}}\mathbf{u}}$ Hermitian, and commuting partial derivatives make $\mathbf{F}_{\bar{\mathbf{u}}} = E_{\bar{\mathbf{u}}\bar{\mathbf{u}}}$ symmetric. A concrete case is a lattice with a Kerr nonlinearity,

$$\mathbf{F} = \mathbf{K}\mathbf{u} + g|\mathbf{u}|^2 \odot \mathbf{u}, \quad E = \langle \mathbf{u}, \mathbf{K}\mathbf{u} \rangle_{\mathbb{C}} + \frac{g}{2} \sum_i |u_i|^4, \quad (73)$$

with $\mathbf{K}$ Hermitian and g real, for which $\mathbf{F}_{\mathbf{u}} = \mathbf{K} + 2g \text{diag}(|\mathbf{u}|^2)$ and $\mathbf{F}_{\bar{\mathbf{u}}} = g \text{diag}(\mathbf{u}^2)$. This example recovers the exact limit of Near-Equilibrium Propagation (NEP) [30].

Comparison with the trajectory case. A trajectory construction must also regenerate the reversed background. With the conjugation TRM of Sec. 7, the Kerr example requires $\mathbf{K}$ to be real, while tangent matching requires it to be Hermitian; together these conditions make $\mathbf{K}$ real symmetric. With the generalized $\mathcal{PT}$ mirror of Sec. 9.2, the two conditions are $\mathbf{K} = \Pi \bar{\mathbf{K}} \Pi$ and $\mathbf{K} = \mathbf{K}^\dagger$. The stationary construction needs no reversed background and therefore requires only Hermiticity.

The real-valued case. For real fields, $\mathbf{F}_{\bar{\mathbf{u}}}$ is absent and Hermiticity is symmetry, so condition (64) reduces to $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ at the steady state. The adjoint and gradient are then real,

$$\mathbf{F}_{\mathbf{u}}^T \mathbf{a} + \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}^*] = \mathbf{0}, \quad \frac{d\chi}{d\mathbf{p}} = \langle \mathbf{a}, \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}^*, \mathbf{p}_F] - \mathbf{f}_{\mathbf{p}_f}(\mathbf{p}_f) \rangle, \quad (74)$$

with $\psi_{\mathbf{u}}$ the ordinary real gradient; the factor 2Re disappears together with the Wirtinger convention. The construction is then Equilibrium Propagation [29]; Appendix I.3 derives this static adjoint and relates it to the trajectory adjoint of Sec. 6.

Literature. EP was first introduced by Scellier and Bengio [29], theoretically applied to nonlinear resistive networks by Kendall et al. [88] and experimentally implemented on an Ising machine by Laydevant et al. [74]. It was later extended by Laborieux and Zenke [89] to holomorphic networks, where exact gradients can be recovered from finite-amplitude complex oscillations; there it is the nudge amplitude that is complexified and the dynamics must be holomorphic, $\mathbf{F}_{\bar{\mathbf{u}}} = \mathbf{0}$. A broad literature has clarified the relation between EP and recurrent backpropagation [90, 91], reduced finite-nudge bias via symmetric nudging and scaled EP to deep convolutional networks [83], introduced frequency propagation as a related frequency-multiplexed learning rule [92], studied EP in strongly multistable systems [93] and developed approximate extensions beyond reciprocal energy-based dynamics [76, 77]. In particular, Scurria et al. [78] generalize to non-reciprocal systems by modifying the learning dynamics. Lin et al. [25] propose a generalized Equilibrium Propagation that organizes two-phase rules by the perturbative order of the nudge at the free equilibrium, recovering EP at first order and Coupled Learning at second, and apply it to linear resistor networks.

Sajnok and Matuszewski [30] derived a complex-valued equilibrium-propagation rule for nonlinear networks and continuous fields. When their tangent conditions hold exactly, they coincide with Eq. (64), and their update rule agrees with the direct-adjoint construction presented here. Recovering $\mathbf{a}$ in this

way requires Hermiticity of $\mathbf{F}_{\mathbf{u}}$, which linear loss violates. Their NEP formulation treats this violation perturbatively in weakly dissipative systems.

Guillamon et al. [75] experimentally implement the conjugated-adjoint construction, $\mathbf{c} = \bar{\mathbf{a}}$, in a lossy reciprocal network with one localized nonlinear element. A global phase rotation makes its anomalous tangent coefficient real, satisfying conditions (65). Here the symmetry condition on $\mathbf{F}_{\mathbf{u}}$ is preserved by reciprocal linear loss, so the construction remains exact in the infinitesimal-nudge limit despite damping. For several local nonlinear elements, the same phase correction requires their nonzero anomalous coefficients to share a phase modulo π .

Dal Cin et al. [94] recover exact gradients in reciprocal linear systems, including dissipative systems, as the nudge vanishes. Their nonlinear optical extension is perturbative in the nonlinear response.

8.3 Quantum Equilibrium Propagation

Quantum Equilibrium Propagation (QEP) trains the parameters of a quantum Hamiltonian through the response of its equilibrium state [31–33]. We consider the normalized, nondegenerate ground state $\mathbf{u}^*$ of $\mathbf{H}(\mathbf{p}) = \sum_j p_j \mathbf{A}_j$, with real parameters p_j and fixed Hermitian operators $\mathbf{A}_j$. Inputs, when present, enter through a further parameter block held fixed during each training step. Although the Schrödinger equation is linear in the state, the energy expectation depends nonlinearly on the normalized state. QEP is therefore a constrained version of the nonlinear stationary construction in Theorem 8.1.

The outputs are expectation values $y_\ell = \langle \mathbf{u}^*, \mathbf{O}_\ell \mathbf{u}^* \rangle_{\mathbb{C}}$ of Hermitian observables, and the loss is $\chi = \psi(\mathbf{y})$. In the free phase, the device prepares the ground state and measures these outputs. The derivative, $\psi_{\mathbf{y}} = d\psi/d\mathbf{y}$ (with components ψ_{y_ℓ}), typically referred to as the error signal and evaluated from the free-state readout, gives the adjoint source and its Hamiltonian implementation [31]:

$$\begin{aligned} \mathbf{f}_{\text{adj}} &= \chi_{\bar{\mathbf{u}}} = \sum_{\ell} \psi_{y_\ell} \mathbf{O}_\ell \mathbf{u}^*, \\ \mathbf{H}_\epsilon &= \mathbf{H} + \epsilon \sum_{\ell} \psi_{y_\ell} \mathbf{O}_\ell. \end{aligned} \tag{75}$$

Because the loss depends on operator expectation values, the source is the action of a Hermitian operator on the state. It is therefore implemented as a Hamiltonian perturbation, whereas the preceding sections add an external force. The error signal components are held fixed during nudging. The device must support these Hamiltonian terms and prepare the ground state $\mathbf{u}^{*,\epsilon}$ of the nudged Hamiltonian.

Here, state variations must preserve unit norm. We fix the arbitrary global phase so that these variations lie in the tangent space orthogonal to $\mathbf{u}^*$. Appendix I.6 shows that, on this space, the Hermiticity of $\mathbf{H}$ automatically gives both tangent reciprocity conditions of Theorem 8.1. The Hamiltonian nudge supplies the adjoint source projected onto this space, so the infinitesimal state response is the adjoint. A positive excitation gap makes this response unique. The appendix also relates this construction to the Onsager reciprocity used by Wanjura and Marquardt [31].

Writing $\langle \mathbf{A}_j \rangle_\epsilon = \langle \mathbf{u}^{*,\epsilon}, \mathbf{A}_j \mathbf{u}^{*,\epsilon} \rangle_{\mathbb{C}}$, their gradient formula is our stationary adjoint overlap:

$$\frac{d\chi}{dp_j} = \lim_{\epsilon \rightarrow 0} \frac{\langle \mathbf{A}_j \rangle_\epsilon - \langle \mathbf{A}_j \rangle_0}{\epsilon} = 2 \operatorname{Re} \langle \mathbf{a}, \mathbf{A}_j \mathbf{u}^* \rangle_{\mathbb{C}}. \tag{76}$$

The proof is given in Appendix I.6.

QEP obtains this gradient without reading out the adjoint field itself. It measures $\langle \mathbf{A}_j \rangle$ in the free and nudged states and divides their difference by ϵ . This works because an expectation value is quadratic in the state: its first-order change pairs the free state with the state response, which is the adjoint. Since $\mathbf{A}_j = \partial \mathbf{H} / \partial p_j$, that pairing is exactly the overlap needed for the parameter gradient. No phase-resolved field measurement is required. When $\mathbf{A}_j$ describes a local coupling, the readout is local in the sense of Sec. 4; for a two-spin coupling it is the corresponding spin correlator.

At finite ϵ , the one-sided estimate has an $O(\epsilon)$ bias. Opposite nudges reduce this to $O(\epsilon^2)$, while repeated state preparations are needed to estimate the expectation values. The nudge must stay within the linear-response regime and be large enough to resolve the observable differences above measurement noise; Appendix I.6.8 discusses this trade-off.

Literature. Wanjura and Marquardt [31] derive QEP from Onsager reciprocity and demonstrate it numerically for many-body phase recognition, phase-diagram exploration, and sensitivity optimisation. Quantum versions of EP were developed independently by Massar and Mognetti [32] and Scellier [33].

9 General intertwiners and twisted reciprocity

Reciprocity of the tangent operators is a sufficient condition for access to the adjoint on the same hardware. More generally, an intertwiner can map the adjoint operators onto the forward operators. We state the construction for Schrödinger-type systems. The forward-time conjugated Schrödinger adjoint $\mathbf{c}$ of Eq. (42) generally evolves under transposed Wirtinger tangent operators. A linear system needs only an intertwiner $\mathbf{S}$, whereas a nonlinear system also needs a TRM $\mathcal{QK}$ to regenerate the reversed background.

For both constructions below, let τ denote time in the adjoint experiment and set

$$r := T - \tau. \quad (77)$$

Write the distributed and terminal loss derivatives as

$$\mathbf{q}(\tau) := \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(r), \overline{\mathbb{P}\mathbf{u}(r)}], \quad \mathbf{q}_T := \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T), \overline{\mathbb{P}\mathbf{u}(T)}]. \quad (78)$$

Theorem 9.1 (Sufficient conditions for intertwined physical adjoints in Schrödinger-type systems). *Consider the Schrödinger system (41). The adjoint can be obtained on the same hardware in the following cases:*

- **Linear.** Suppose

$$\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t] = \mathbf{K}(\mathbf{p}_K, t)\mathbf{u}$$

and, at the current parameter value, there is a known invertible matrix $\mathbf{S}$ that is fixed over the full time schedule and satisfies

$$\mathbf{S}\mathbf{K}(\mathbf{p}_K, t)^T\mathbf{S}^{-1} = \mathbf{K}(\mathbf{p}_K, t), \quad \forall t \in [0, T]. \quad (79)$$

- **Nonlinear.** Let $\mathcal{K}$ denote complex conjugation and

$$\Theta_{\mathcal{Q}} := \mathcal{QK}, \quad (80)$$

where $\mathcal{Q}$ is an invertible complex-linear map. Suppose $\mathbf{F}$ is continuously Wirtinger differentiable on an operating domain invariant under $\mathbf{u} \mapsto \mathcal{Q}\bar{\mathbf{u}}$ and obeys

$$\mathbf{F}[\mathcal{Q}\bar{\mathbf{u}}, \overline{\mathcal{Q}\mathbf{u}}, \mathbf{p}_F, t] = \mathcal{Q}\overline{\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t]}, \quad \forall t \in [0, T], \quad (81)$$

throughout that domain. Suppose, independently, that an invertible matrix $\mathbf{S}$ satisfies, along the forward trajectory,

$$\mathbf{S}\mathbf{F}_{\mathbf{u}}^T\mathbf{S}^{-1} = \overline{\mathbf{F}_{\mathbf{u}}}, \quad (82)$$

$$\mathbf{S}\overline{\mathbf{F}_{\bar{\mathbf{u}}}}^T\mathbf{S}^{-1} = \overline{\mathbf{F}_{\bar{\mathbf{u}}}}, \quad \forall t \in [0, T], \quad (83)$$

where both Jacobians are evaluated at $[\mathbf{u}(t), \overline{\mathbf{u}(t)}, \mathbf{p}_F, t]$.

Other equation classes. For real linear first-order systems, the same similarity condition gives $\mathbf{d} = \mathbf{S}\mathbf{a}$. Real linear second-order systems also require $\mathbf{S}\mathbf{M}^T\mathbf{S}^{-1} = \mathbf{M}$ and $\mathbf{S}\mathbf{D}^T\mathbf{S}^{-1} = \mathbf{D}$. For example, an additional gyroscopic term $\mathbf{G}\dot{\mathbf{u}}$, with $\mathbf{G}^T = -\mathbf{G}$, is covered by $\mathbf{S} = \mathbf{\Pi}$ if a constant real mirror $\mathbf{\Pi} = \mathbf{\Pi}^T = \mathbf{\Pi}^{-1}$ preserves the symmetric mass, damping and stiffness operators and satisfies $\mathbf{\Pi}\mathbf{G}\mathbf{\Pi} = -\mathbf{G}$. Then $\mathbf{\Pi}(\mathbf{D} + \mathbf{G})^T\mathbf{\Pi} = \mathbf{D} + \mathbf{G}$. Appendix G treats the nonlinear, undamped case, where a $\mathcal{PT}$ -TRM must also regenerate the background. Nonlinear real first-order dynamics have no generic unchanged-device reverse trajectory.² Stationary systems need no TRM or schedule reversal: the same linear intertwining condition applies, while nonlinear nudging requires a stable operating point and continuation on the same solution branch.

9.1 Linear Schrödinger systems: spatial symmetries and twisted reciprocity

Using the notation above, the complex conjugate of the linear Schrödinger adjoint is

$$i\dot{\mathbf{c}}(\tau) + \mathbf{K}(\mathbf{p}_K, r)^T \mathbf{c}(\tau) + \mathbf{q}(\tau) = \mathbf{0}, \quad \mathbf{c}(0) = i\mathbf{q}_T. \quad (84)$$

Set $\mathbf{d} = \mathbf{S}\mathbf{c}$ and multiply this equation by $\mathbf{S}$. Condition (79) gives

$$i\dot{\mathbf{d}}(\tau) + \mathbf{K}(\mathbf{p}_K, r)\mathbf{d}(\tau) + \mathbf{S}\mathbf{q}(\tau) = \mathbf{0}, \quad \mathbf{d}(0) = i\mathbf{S}\mathbf{q}_T. \quad (85)$$

This has the same dynamical operator as the forward system and proves the linear part of Theorem 9.1. $\square$

Which $\mathbf{S}$ qualify, and at what cost? At each parameter point, a known $\mathbf{S}(\mathbf{p}_K)$ must satisfy Eq. (79) at all times during that experiment. A common $\mathbf{S}$ for the full trainable family avoids recalibration. Otherwise $\mathbf{S}(\mathbf{p}_K)$ must be recalibrated after every update. A signed permutation relocates ports and may add sign flips. A unitary $\mathbf{S}$ preserves amplitudes but can require coherent source combinations. A diagonal $\mathbf{S}$ preserves source support but rescales its amplitude. A general invertible $\mathbf{S}$ requires weighted source synthesis and inverse weighting during recovery; its condition number sets the required dynamic range. Parameter updates must remain within a family for which such a transformation is available.

Physical implementation. The experimenter must be able to apply the transformed drive and initial condition and measure the complex field $\mathbf{d}$. The experimenter applies the drive $-\mathbf{S}\mathbf{q}$, prepares $\mathbf{d}(0) = i\mathbf{S}\mathbf{q}_T$, and recovers the conjugate of the adjoint digitally as

$$\mathbf{c} = \mathbf{S}^{-1}\mathbf{d}. \quad (86)$$

This is one additional finite-amplitude run. The transformation does not appear in the forward equation. If $\mathbf{S}$ is a spatial symmetry, these operations relocate or coherently combine source and measurement ports.

Reciprocal systems and fully forward mode training. For reciprocal hardware, $\mathbf{K}^T = \mathbf{K}$, the identity $\mathbf{S} = \mathbf{I}$ satisfies Eq. (79). This recovers the linear result of Theorem 7.1. Suppose the hardware also has a spatial symmetry

$$\mathbf{S}\mathbf{K}(t)\mathbf{S}^{-1} = \mathbf{K}(t), \quad \mathbf{S}\mathbf{P}\mathbf{S}^{-1} = \mathbb{P}^{\text{in}},$$

which maps the output DOFs onto the input DOFs. The same $\mathbf{S}$ is then an intertwiner, so the adjoint source and initial condition can be injected where the data enter. This is fully forward mode training [43]. Reciprocity supplies the transpose propagation, and the spatial symmetry routes the adjoint excitation. The parameter updates must preserve that symmetry. The original derivation uses Green's functions for static Helmholtz propagation; Eq. (85) extends the construction to time-dependent Schrödinger trajectories.

²Here we exclude prescribed trajectory compensation; see the paragraphs on [damping compensation](#) in Sec. 5.2 and [trajectory compensation](#) in Sec. 6.2.

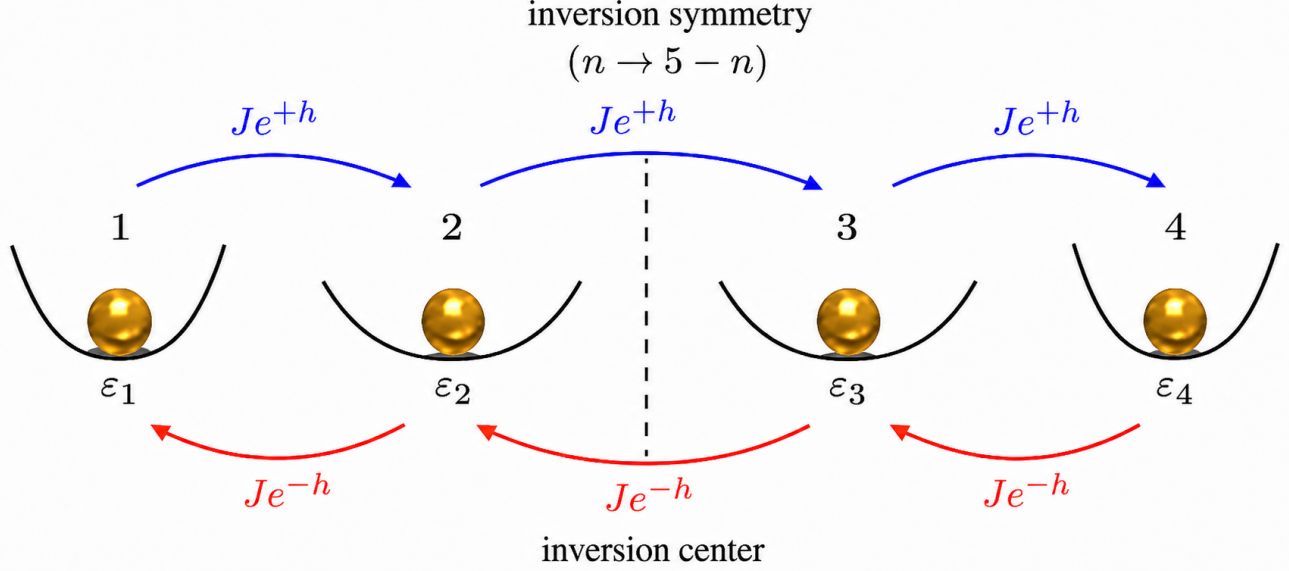


Figure 3: The Hatano–Nelson chain of Eq. (88) for $N = 4$. The rightward hopping is amplified by e^h and the leftward hopping is attenuated by e^{-h} , so $\mathbf{K}^T \neq \mathbf{K}$. Transposition reverses every hopping arrow, and spatial inversion reverses it again. For a mirror-symmetric on-site potential, $\varepsilon_n = \varepsilon_{N+1-n}$, the chain obeys $\mathbf{S}\mathbf{K}^T\mathbf{S}^{-1} = \mathbf{K}$. The transformed adjoint therefore propagates on the same chain with the same bias.

Non-reciprocal systems: twisted reciprocity. We call a non-reciprocal system *twisted reciprocal* if an orthogonal transformation $\mathbf{S}$ satisfies Eq. (79). Write $\mathbf{K} = \mathbf{K}_s + \mathbf{K}_a$, where $\mathbf{K}_s^T = \mathbf{K}_s$ and $\mathbf{K}_a^T = -\mathbf{K}_a$. Then Eq. (79) is equivalent to

$$\mathbf{S}\mathbf{K}_s\mathbf{S}^{-1} = \mathbf{K}_s, \quad \mathbf{S}\mathbf{K}_a\mathbf{S}^{-1} = -\mathbf{K}_a. \quad (87)$$

The reciprocal part is even under the symmetry, while the non-reciprocal part is odd. Examples include a hopping bias [42, 95, 96], a direction of rotation [97–99], or an in-plane magnetization. In the last case the symmetry supplies the field reversal in the Onsager–Casimir relation $\mathbf{K}(\mathbf{B})^T = \mathbf{K}(-\mathbf{B})$ [100]. The operator remains non-reciprocal, $\mathbf{K}^T \neq \mathbf{K}$. Equation (79) is related to the transpose-type symmetries in the non-Hermitian symmetry classification [101], used here as a trainability condition.

Example of twisted reciprocity: a Hatano–Nelson chain. Consider N sites on an open chain with asymmetric nearest-neighbor hoppings and a real on-site potential [42],

$$K_{n+1,n} = J e^{+h}, \quad K_{n,n+1} = J e^{-h}, \quad K_{nn} = \varepsilon_n, \quad J, h, \varepsilon_n \in \mathbb{R}. \quad (88)$$

For $N = 4$, the operator and spatial inversion $S_{mn} = \delta_{m,N+1-n}$ are

$$\mathbf{K} = \begin{bmatrix} \varepsilon_1 & J e^{-h} & 0 & 0 \\ J e^{+h} & \varepsilon_2 & J e^{-h} & 0 \\ 0 & J e^{+h} & \varepsilon_3 & J e^{-h} \\ 0 & 0 & J e^{+h} & \varepsilon_4 \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (89)$$

The chain has an exponentially biased directional response and is a standard model of the non-Hermitian skin effect [101]. Related platforms include feedback-biased active metamaterials [102, 103], non-reciprocal

topoelectrical circuits [95], non-reciprocal photonic lattices [96], and acoustic circulators [98]. The untransformed adjoint in Eq. (84) propagates under $\mathbf{K}^T$, which the hardware does not provide.

Spatial inversion gives Eq. (79) when $\varepsilon_n = \varepsilon_{N+1-n}$. The transformed field is obtained in one additional finite-amplitude run with the same bias h . If the loss is measured at the right edge, the source $-\mathbf{S}\mathbf{q}$ and initial condition $i\mathbf{S}\mathbf{q}_T$ are applied at the left edge. Since $\mathbf{S}^{-1} = \mathbf{S}$, the adjoint is $\mathbf{c} = \mathbf{S}\mathbf{d}$. Heterogeneous profiles are allowed when $J_n = J_{N-n}$, $h_n = h_{N-n}$, and $\varepsilon_n = \varepsilon_{N+1-n}$. These parameter ties must be preserved during training.

Literature. Fully Forward Mode training [43] is an example of the reciprocal spatial-symmetry construction above. For linear resistor networks, Lin et al. [25] obtain exact gradients from the weighted-reciprocity identity $\mathbf{R}\boldsymbol{\Omega}^T\mathbf{R}^{-1} = \boldsymbol{\Omega}$, with $\mathbf{S} = \mathbf{R}$, accessible through voltage- and current-mode measurements. To the best of the authors' knowledge, prior work has not exploited orthogonal transpose intertwiners to generate physical adjoints on the same non-reciprocal hardware.

9.2 Nonlinear Schrödinger systems: time-reversal mirrors and intertwined reciprocity

The nonlinear part of Theorem 9.1 is proved in Appendix F.4.

Physical adjoint algorithm. The TRM generates the reversed background

$$\mathbf{w}(\tau) = \overline{Q\mathbf{u}(r)}. \quad (90)$$

The experimenter must be able to apply the transformed drive and initial condition and measure the resulting complex field. The transformations Q and $\mathbf{S}$ are known and constant during each experiment. If either depends on the parameters, it must be recalibrated after an update. Run a nudged trajectory $\mathbf{w}^\epsilon$ on the same hardware, with the explicit schedule evaluated at r and the applied drive

$$\mathbf{f}^\epsilon(\tau) = \overline{Q\mathbf{f}(r, \mathbf{p}_f)} - \epsilon Q\mathbf{S}\mathbf{q}(\tau). \quad (91)$$

The initial condition is

$$\mathbf{w}^\epsilon(0) = \overline{Q\mathbf{u}(T)} + \epsilon i Q\mathbf{S}\mathbf{q}_T. \quad (92)$$

The differential response gives the transformed adjoint,

$$\mathbf{d}(\tau) := \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(\tau) - \mathbf{w}(\tau)}{\epsilon} = Q\mathbf{S}\mathbf{c}(\tau), \quad \mathbf{c}(\tau) = \mathbf{S}^{-1}Q^{-1}\mathbf{d}(\tau). \quad (93)$$

This construction is exact in the infinitesimal-nudge limit.

Example: $\mathcal{PT}$ symmetry. Let Π be a real signed-permutation parity involution. Parity enters through the TRM, while no additional tangent intertwiner is used:

$$Q = \Pi, \quad \mathbf{S} = \mathbf{I}, \quad \mathbf{d} = Q\mathbf{S}\mathbf{c} = \Pi\mathbf{c}. \quad (94)$$

Condition (81) is then the $\mathcal{PT}$ covariance

$$\mathbf{F}[\Pi\bar{\mathbf{u}}, \Pi\mathbf{u}, \mathbf{p}_F, t] = \overline{\Pi\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t]}. \quad (95)$$

The direct matching conditions between the physical tangent at the reversed background and the adjoint tangent are

$$\Pi\mathbf{F}_{\mathbf{u}}[\Pi\bar{\mathbf{u}}, \Pi\mathbf{u}, \mathbf{p}_F, t]\Pi = \mathbf{F}_{\mathbf{u}}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t]^T, \quad (96)$$

$$\Pi\mathbf{F}_{\bar{\mathbf{u}}}[\Pi\bar{\mathbf{u}}, \Pi\mathbf{u}, \mathbf{p}_F, t]\Pi = \overline{\mathbf{F}_{\bar{\mathbf{u}}}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F, t]^T}. \quad (97)$$

Differentiating the $\mathcal{PT}$ covariance shows that the left-hand sides are $\overline{\mathbf{F}}_{\mathbf{u}}$ and $\overline{\mathbf{F}}_{\overline{\mathbf{u}}}$, respectively. Equations (96) and (97) therefore reduce to conditions (82) and (83) with $\mathbf{S} = \mathbf{I}$. The covariance regenerates the background but does not by itself impose these tangent conditions.

For the Kerr force in Eq. (56), the TRM condition requires $\mathbf{K} = \mathbf{\Pi}\mathbf{K}\mathbf{\Pi}$, while tangent matching requires Hermiticity, i.e. $\mathbf{K} = \mathbf{K}^T$. The local Kerr terms satisfy both tangent conditions. The $\mathcal{PT}$ mirror thus permits a Hermitian coupling matrix whose complex coupling phases are reversed by the spatial inversion $\mathbf{\Pi}$. In contrast, with ordinary conjugation and $\mathbf{S} = \mathbf{I}$, background reversal requires $\mathbf{K} = \overline{\mathbf{K}}$, while tangent matching requires $\mathbf{K}^T = \mathbf{K}$; together they restrict $\mathbf{K}$ to be real symmetric. The recovered adjoint is $\mathbf{c} = \mathbf{\Pi}\mathbf{d}$.

Example: a nonlinear Hatano–Nelson chain. Consider the open chain

$$K_{n+1,n} = J_n e^{h_n}, \quad K_{n,n+1} = J_n e^{-h_n}, \quad K_{nn} = \varepsilon_n, \quad F_n = (\mathbf{K}\mathbf{u})_n + g_n |u_n|^2 u_n, \quad (98)$$

with $J_n, h_n, \varepsilon_n, g_n \in \mathbb{R}$. Ordinary conjugation regenerates the nonlinear background, so $\mathcal{Q} = \mathbf{I}$. Define

$$\mathbf{S}_{\text{HN}} := \text{diag}(s_1, \dots, s_N), \quad \frac{s_{n+1}}{s_n} = e^{2h_n}, \quad s_1 > 0. \quad (99)$$

The spatial inversion used above routes sources without amplitude weighting but requires mirror-tied coefficients. In contrast, the diagonal choice satisfies $\mathbf{S}_{\text{HN}} \mathbf{K}^T \mathbf{S}_{\text{HN}}^{-1} = \mathbf{K}$ for arbitrary open-chain profiles and commutes with the local Kerr terms, at the cost of site-dependent amplitudes. The Wirtinger Jacobians are

$$\mathbf{F}_{\mathbf{u}} = \mathbf{K} + 2 \text{diag}(g_n |u_n|^2), \quad \mathbf{F}_{\overline{\mathbf{u}}} = \text{diag}(g_n u_n^2). \quad (100)$$

The diagonal nonlinear terms commute with $\mathbf{S}_{\text{HN}}$, so Eqs. (82) and (83) hold with $\mathbf{S} = \mathbf{S}_{\text{HN}}$. The measured perturbation is $\mathbf{d} = \mathbf{S}_{\text{HN}} \mathbf{c}$.

The diagonal intertwiner preserves the support of the nudge but rescales its amplitude at each site. The adjoint is recovered as $\mathbf{c} = \mathbf{S}_{\text{HN}}^{-1} \mathbf{d}$. If the hopping biases are trained, $\mathbf{S}_{\text{HN}}$ must be recalibrated after every update.

Literature. Dal Cin et al. [94] use an approximate intertwining relation for the scattering matrix linearized about a nonlinear optical steady state; their two-measurement gradient is perturbative in both the nonlinear response and the nudge amplitude. In concurrent numerical work, Li et al. [79] combine a $\mathcal{PT}$ mirror with prescribed gain–loss compensation in nonlinear resonator chains, analogous to the [damping compensation in Sec. 5.2](#). Their gradients are exact in the infinitesimal-nudge limit.

9.3 Onsager reciprocity

Onsager reciprocity [104, 105] is the special case in which the intertwiner comes from a physical mobility. For real state variables, let the constant mobility satisfy $\mathbf{V} = \mathbf{V}^T \succ 0$ and define the Onsager inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{V}^{-1}} = \mathbf{x}^T \mathbf{V}^{-1} \mathbf{y}. \quad (101)$$

An operator $\mathbf{K}$ is $\mathbf{V}$ -reciprocal if it is self-adjoint in this metric,

$$\mathbf{V} \mathbf{K}^T \mathbf{V}^{-1} = \mathbf{K}. \quad (102)$$

Equivalently, $\mathbf{K} = \mathbf{V} \mathbf{H}$ with $\mathbf{H} = \mathbf{H}^T$. The same transpose identity gives $\mathbf{S} = \mathbf{V}$ for linear Schrödinger systems, including complex symmetric $\mathbf{H}$. Here $\mathbf{V}$ is part of the dynamics. If the physical source enters as $\mathbf{f} = \mathbf{V} \mathbf{j}$, setting $\mathbf{j} = -\mathbf{q}$ produces the required applied drive $-\mathbf{V} \mathbf{q}$. With $\mathbf{d} = \mathbf{V} \mathbf{c}$, the adjoint is recovered as $\mathbf{c} = \mathbf{V}^{-1} \mathbf{d}$. If the trainable parameters enter $\mathbf{H}$ and $\mathbf{V}$ is fixed, the weighting also cancels from the gradient overlap,

$$2 \text{Re} [\mathbf{c}^T \mathbf{K}_{\mathbf{p}} \mathbf{u}] = 2 \text{Re} [\mathbf{d}^T \mathbf{H}_{\mathbf{p}} \mathbf{u}]. \quad (103)$$

Natural source variables can therefore supply the Onsager weighting, and the overlap above does not require explicit inverse weighting. For other parameter blocks, this cancellation need not hold. Symmetry and positivity distinguish an Onsager mobility from a general intertwiner; neither is required by the linear construction.

Nonlinear systems. A constant mobility also supplies a nonlinear tangent intertwiner. Let $E[\mathbf{u}, \bar{\mathbf{u}}] \in \mathbb{R}$ be twice continuously differentiable and let $\mathbf{V} = \mathbf{V}^\dagger \succ 0$ be constant. If

$$\mathbf{F} = \mathbf{V}E_{\bar{\mathbf{u}}}, \quad (104)$$

then $\mathbf{F}_{\mathbf{u}} = \mathbf{V}E_{\bar{\mathbf{u}}\mathbf{u}}$ and $\mathbf{F}_{\bar{\mathbf{u}}} = \mathbf{V}E_{\bar{\mathbf{u}}\bar{\mathbf{u}}}$. Reality of E makes $E_{\bar{\mathbf{u}}\mathbf{u}}$ Hermitian, while commuting Wirtinger derivatives make $E_{\bar{\mathbf{u}}\bar{\mathbf{u}}}$ symmetric. Hence

$$\overline{\mathbf{V}}\mathbf{F}_{\mathbf{u}}^T\overline{\mathbf{V}}^{-1} = \overline{\mathbf{F}_{\mathbf{u}}}, \quad (105)$$

$$\overline{\mathbf{V}}\mathbf{F}_{\bar{\mathbf{u}}}^T\overline{\mathbf{V}}^{-1} = \overline{\mathbf{F}_{\bar{\mathbf{u}}}}, \quad (106)$$

so Eqs. (82) and (83) hold with

$$\mathbf{S} = \overline{\mathbf{V}}. \quad (107)$$

For a real symmetric mobility, $\mathbf{S} = \mathbf{V}$. This supplies the tangent intertwiner; the TRM condition (81) remains independent. The physical perturbation and adjoint recovery are given by Eq. (93). Constancy of $\mathbf{V}$ is essential here, but positivity is not needed for the intertwining identities.

The nonlinear Hatano–Nelson chain (98) has the form (104) with $\mathbf{V} = \mathbf{S}_{\text{HN}}$, while ordinary conjugation independently supplies its TRM.

For real stationary gradient flows, $\mathbf{F} = \mathbf{V}E_{\mathbf{u}}$ with constant $\mathbf{V} = \mathbf{V}^T \succ 0$, the nudge is $\epsilon \mathbf{V}\mathbb{P}\psi_{\mathbf{u}}$ and the adjoint is

$$\mathbf{a} = \mathbf{V}^{-1} \lim_{\epsilon \rightarrow 0} \frac{\mathbf{u}^{*\epsilon} - \mathbf{u}^*}{\epsilon}. \quad (108)$$

Positive definiteness makes both free and nudged relaxation descend the tilted energy in the $\mathbf{V}^{-1}$ metric. A stable operating point and continuation on the same solution branch are still required.

Appendices D.1 and D.2 derive the real second- and first-order constructions. Appendix D.2.1 gives the RC-network realization, where current and charge ports supply the weighting.

10 Conclusions

We have unified a broad range of formally exact physical gradient methods [20–24, 26, 29, 34–38, 43] using the adjoint method from PDE-constrained optimization. We have clarified that linear and nonlinear systems behave fundamentally differently and are governed by different sufficient conditions for obtaining the adjoint field on the same hardware. For linear systems, reciprocity is sufficient and the adjoint follows from a single finite-amplitude experiment, whereas nonlinear trajectory systems additionally require a time-reversal mirror, reciprocity of the linearized dynamics, and infinitesimal nudging. For stationary problems, the framework covers Equilibrium Propagation, Quantum Equilibrium Propagation, and in situ gradients in Helmholtz-type systems.

We have further shown that reciprocity is only the simplest instance of more general intertwining conditions. For linear systems, it suffices that the forward operator and its transpose are related by a known, invertible transformation that is constant during each experiment. This recovers fully forward mode training through spatial symmetries and permits exact physical gradients in a class of non-Hermitian, non-reciprocal systems. For nonlinear trajectories, the construction also requires a generalized time-reversal mirror to regenerate the reversed background and an independent intertwining condition that matches the tangent dynamics to the adjoint. The gradient is then obtained in the infinitesimal-nudge limit. Examples include $\mathcal{PT}$ -symmetric systems and linear and nonlinear Hatano–Nelson chains [42]. Onsager reciprocity is

the special case in which a constant physical mobility supplies the intertwiner, for both linear and nonlinear dynamics. It also covers stationary gradient flows near stable fixed points, where no time-reversal mirror is needed. The constructions allow explicit time dependence, with schedules replayed in reverse for trajectory adjoints.

Because on-device gradients are far easier to obtain in linear than in nonlinear hardware, parameter encoding (structural nonlinearity) [59, 60] and source-position encoding [58] are particularly attractive routes to physical computing: the device computes nonlinearly in its inputs, yet its adjoint field, and with it the gradient, remains accessible at finite amplitude and with damping present [36, 106]. The non-autonomous generalization makes the temporal modulation of the internal forces itself a trainable resource and may thereby enable novel ways of computing with physical systems.

As discussed in Sec. 4, when the parameters enter the forward operators linearly, the gradient overlaps are assembled entirely from measured trajectories and never reference the parameter values, so slow drift, aging, and calibration offsets are absorbed automatically. What must still be known is the structure of the model: which measured signals to pair, and where the trainable parameters sit in the equations of motion. How accurately that structure must be known and how the gradients degrade when it is wrong remain open questions.

Remark on concurrent work. Shortly after our manuscript appeared on arXiv, three works addressing specific cases of physical backpropagation covered by our framework were posted: nonlinear stationary wave scattering with a single nonlinear element by Guillaumon et al. [75], $\mathcal{PT}$ -assisted trajectory learning in nonlinear resonator chains by Li et al. [79], and time-domain adjoint experiments in linear RLC circuits by Kwon et al. [71]. We discuss their relation to our results in the corresponding sections.

Acknowledgments We thank Dr. Sam Dillavou, Dr. Fatih Dinc and Andy Davis for helpful discussions and feedback on the manuscript. C.B. was supported by the Swiss National Science Foundation (SNSF) through a Postdoc.Mobility fellowship (P500PT 217673/1). YG and HT were sponsored by AFOSR MURI Grant No. FA955022-1-0203 and by the Army Research Office under Grant No. W911NF-25-1-0261. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-2210452.

Data availability Notebooks implementing representative physical adjoint algorithms and numerical gradient checks are publicly available in Ref. [57].

References

- [1] Herbert Jaeger, Beatriz Noheda, and Wilfred G Van Der Wiel. Toward a formal theory for computing machines made out of whatever physics offers. *Nature Communications*, 14(1):4911, 2023.
- [2] Tyler W Hughes, Ian AD Williamson, Momchil Minkov, and Shanhui Fan. Wave physics as an analog recurrent neural network. *Science Advances*, 5(12):eaay6946, 2019.
- [3] Farzad Zangeneh-Nejad, Dimitrios L Sounas, Andrea Alù, and Romain Fleury. Analogue computing with metamaterials. *Nature Reviews Materials*, 6(3):207–225, 2021.
- [4] Tena Dubček, Daniel Moreno-Garcia, Thomas Haag, Parisa Omidvar, Henrik R Thomsen, Theodor S Becker, Lars Gebraad, Christoph Bärlocher, Fredrik Andersson, Sebastian D Huber, et al. In-sensor passive speech classification with phononic metamaterials. *Advanced Functional Materials*, 34(17):2311877, 2024.

- [5] Sam Dillavou, Menachem Stern, Andrea J Liu, and Douglas J Durian. Demonstration of decentralized physics-driven learning. *Physical Review Applied*, 18(1):014040, 2022.
- [6] Abu Sebastian, Manuel Le Gallo, Riduan Khaddam-Aljameh, and Evangelos Eleftheriou. Memory devices and applications for in-memory computing. *Nature Nanotechnology*, 15(7):529–544, 2020.
- [7] Danijela Marković, Alice Mizrahi, Damien Querlioz, and Julie Grollier. Physics for neuromorphic computing. *Nature Reviews Physics*, 2(9):499–510, 2020.
- [8] Bhavin J Shastri, Alexander N Tait, Thomas Ferreira de Lima, Wolfram HP Pernice, Harish Bhaskaran, C David Wright, and Paul R Prucnal. Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics*, 15(2):102–114, 2021.
- [9] Uğur Teğın, Mustafa Yıldırım, İlker Oğuz, Christophe Moser, and Demetri Psaltis. Scalable optical learning operator. *Nature Computational Science*, 1(8):542–549, 2021.
- [10] Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David AB Miller, and Demetri Psaltis. Inference in artificial intelligence with deep optics and photonics. *Nature*, 588(7836):39–47, 2020.
- [11] Xing Lin, Yair Rivenson, Nezih T Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008, 2018.
- [12] Peter L McMahon. The physics of optical computing. *Nature Reviews Physics*, 5(12):717–734, 2023.
- [13] Tatsuhiro Onodera, Martin M Stein, Benjamin A Ash, Mandar M Sohoni, Melissa Bosch, Ryotatsu Yanagimoto, Marc Jankowski, Timothy P McKenna, Tianyu Wang, Gennady Shvets, Maxim R Shcherbakov, Logan G Wright, and Peter L McMahon. Arbitrary control over multimode wave propagation for machine learning. *Nature Physics*, 22:164–171, 2026.
- [14] Ryotatsu Yanagimoto, Benjamin A Ash, Mandar M Sohoni, Martin M Stein, Yiqi Zhao, Federico Presutti, Marc Jankowski, Logan G Wright, Tatsuhiro Onodera, and Peter L McMahon. Programmable on-chip nonlinear photonics. *Nature*, 649(8096):330–337, 2026.
- [15] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123, 2019.
- [16] Ali Momeni, Babak Rahmani, Benjamin Scellier, Logan G Wright, Peter L McMahon, Clara C Wanjura, Yuhang Li, Anas Skalli, Natalia G Berloff, Tatsuhiro Onodera, et al. Training of physical neural networks. *Nature*, 645(8079):53–61, 2025.
- [17] Vahid Nikkhah, Ali Pirmoradi, Farshid Ashtiani, Brian Edwards, Firooz Aflatouni, and Nader Engheta. Inverse-designed low-index-contrast structures on a silicon photonics platform for vector-matrix multiplication. *Nature Photonics*, 18(5):501–508, 2024.
- [18] Logan G Wright, Tatsuhiro Onodera, Martin M Stein, Tianyu Wang, Darren T Schachter, Zoey Hu, and Peter L McMahon. Deep physical neural networks trained with backpropagation. *Nature*, 601(7894):549–555, 2022.
- [19] İlker Oguz, Niyazi Ulas Dinc, Mustafa Yildirim, Junjie Ke, Innfarn Yoo, Qifei Wang, Feng Yang, Christophe Moser, and Demetri Psaltis. Optical diffusion models for image generation. *Advances in Neural Information Processing Systems*, 37:59150–59173, 2024.

- [20] Michiel Hermans, Michaël Burm, Thomas Van Vaerenbergh, Joni Dambre, and Peter Bienstman. Trainable hardware for dynamical computing using error backpropagation through physical media. *Nature Communications*, 6(1):6729, 2015.
- [21] Tiankuang Zhou, Lu Fang, Tao Yan, Jiamin Wu, Yipeng Li, Jingtao Fan, Huaqiang Wu, Xing Lin, and Qionghai Dai. In situ optical backpropagation training of diffractive optical neural networks. *Photonics Research*, 8(6):940–953, 2020.
- [22] Chaitanya K Mididoddi, Robert J Kilpatrick, Christina Sharp, Philipp Del Hougne, Simon AR Horsley, and David B Phillips. Threading light through dynamic complex media. *Nature Photonics*, 19(4):434–440, 2025.
- [23] Sunil Pai, Zhanghao Sun, Tyler W Hughes, Taewon Park, Ben Bartlett, Ian AD Williamson, Momchil Minkov, Mazyar Milanizadeh, Nathnael Abebe, Francesco Morichetti, et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science*, 380(6643):398–404, 2023.
- [24] Shuaifeng Li and Xiaoming Mao. Training all-mechanical neural networks for task learning through in situ backpropagation. *Nature Communications*, 15(1):10528, 2024.
- [25] Jonathan Lin, Aman Desai, Frank Barrows, and Francesco Caravelli. How to train your resistive network: Generalized equilibrium propagation and analytical learning. *arXiv preprint arXiv:2602.03546*, 2026.
- [26] Víctor López-Pastor and Florian Marquardt. Self-learning machines based on Hamiltonian echo backpropagation. *Physical Review X*, 13(3):031020, 2023.
- [27] Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. Routledge, 2018.
- [28] Andreas Fichtner. *Full seismic waveform modelling and inversion*. Springer Science & Business Media, 2010.
- [29] Benjamin Scellier and Yoshua Bengio. Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Frontiers in Computational Neuroscience*, 11:24, 2017.
- [30] Karol Sajnok and Michał Matuszewski. Near-equilibrium propagation training in nonlinear wave systems. *arXiv preprint arXiv:2510.16084*, 2025.
- [31] C. C. Wanjura and F. Marquardt. Quantum equilibrium propagation for efficient training of quantum systems based on onsager reciprocity. *Nature Communications*, 16:6595, 2025.
- [32] Serge Massar and Bortolo M. Mognetti. Equilibrium propagation: the quantum and the thermal cases. *arXiv preprint arXiv:2405.08467*, 2024.
- [33] Benjamin Scellier. Quantum equilibrium propagation: gradient-descent training of quantum systems. *arXiv preprint arXiv:2406.00879*, 2024.
- [34] Guillaume Pourcel and Maxence Ernoult. Learning long range dependencies through time reversal symmetry breaking. *arXiv preprint arXiv:2506.05259*, 2025.
- [35] Tyler W Hughes, Momchil Minkov, Yu Shi, and Shanhui Fan. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica*, 5(7):864–871, 2018.
- [36] Clara C Wanjura and Florian Marquardt. Fully nonlinear neuromorphic computing with linear wave scattering. *Nature Physics*, 20(9):1434–1440, 2024.

- [37] John Guillaumon, Cheng-Zhen Wang, Zin Lin, and Tsampikos Kottos. In-situ physical adjoint computing in multiple-scattering electromagnetic environments for wave control. *Nature Communications*, 16(1):11466, 2025.
- [38] Shuaifeng Li and Xiaoming Mao. Topological mechanical neural networks as classifiers through in situ backpropagation learning. *arXiv preprint arXiv:2503.07796*, 2025.
- [39] Kelvin Wagner and Demetri Psaltis. Multilayer optical learning networks. *Applied Optics*, 26(23):5061–5076, 1987.
- [40] Xianxin Guo, Thomas D Barrett, Zhiming M Wang, and Alexander I Lvovsky. Backpropagation through nonlinear units for the all-optical training of neural networks. *Photonics Research*, 9(3):B71–B80, 2021.
- [41] James Spall, Xianxin Guo, and Alexander I Lvovsky. Training neural networks with end-to-end optical backpropagation. *Advanced Photonics*, 7(1):016004, 2025.
- [42] Naomichi Hatano and David R Nelson. Localization transitions in non-Hermitian quantum mechanics. *Physical Review Letters*, 77(3):570–573, 1996.
- [43] Zhiwei Xue, Tiankuang Zhou, Zhihao Xu, Shaoliang Yu, Qionghai Dai, and Lu Fang. Fully forward mode training for optical neural networks. *Nature*, 632(8024):280–286, 2024.
- [44] Menachem Stern, Daniel Hexner, Jason W Rocks, and Andrea J Liu. Supervised learning in physical networks: From machine learning to learning machines. *Physical Review X*, 11(2):021045, 2021.
- [45] Menachem Stern and Arvind Murugan. Learning without neurons in physical systems. *Annual Review of Condensed Matter Physics*, 14(1):417–441, 2023.
- [46] Sam Dillavou, Benjamin D Beyer, Menachem Stern, Andrea J Liu, Marc Z Miskin, and Douglas J Durian. Machine learning without a processor: Emergent learning in a nonlinear analog network. *Proceedings of the National Academy of Sciences*, 121(28):e2319718121, 2024.
- [47] Menachem Stern, Sam Dillavou, Dinesh Jayaraman, Douglas J Durian, and Andrea J Liu. Training self-learning circuits for power-efficient solutions. *APL Machine Learning*, 2(1):016114, 2024.
- [48] Joshua A McGinnis, Xinbo Li, and Yoichiro Mori. Coercivity and local convergence of physical learning in linear circuits. *arXiv preprint arXiv:2606.15443*, 2026.
- [49] Junwei Cheng, Chaoran Huang, Jialong Zhang, Bo Wu, Wenkai Zhang, Xinyu Liu, Jiahui Zhang, Yiyi Tang, Hailong Zhou, Qiming Zhang, Min Gu, Jianji Dong, and Xinliang Zhang. Multimodal deep learning using on-chip diffractive optics with in situ training capability. *Nature Communications*, 15(1):6189, 2024.
- [50] Daan de Bos and Marc Serra-Garcia. Multimodal oscillator networks learn to solve a classification problem. *npj Metamaterials*, 2(1):3, 2026.
- [51] Menachem Stern, Adam G Frim, Raúl Candás, Andrea J Liu, and Vijay Balasubramanian. Contrastive learning in tunable dynamical systems. *arXiv preprint arXiv:2603.26969*, 2026.
- [52] Ilker Oguz, Junjie Ke, Qifei Weng, Feng Yang, Mustafa Yildirim, Niyazi Ulas Dinc, Jih-Liang Hsieh, Christophe Moser, and Demetri Psaltis. Forward-forward training of an optical neural network. *Optics Letters*, 48(20):5249–5252, 2023.
- [53] Ali Momeni, Babak Rahmani, Matthieu Malléjac, Philipp Del Hougne, and Romain Fleury. Backpropagation-free training of deep physical neural networks. *Science*, 382(6676):1297–1303, 2023.

- [54] Hao Wang, Ziao Wang, Xiangpeng Liang, Han Zhao, Jianqi Hu, Junjie Jiang, Xing Fu, Jianshi Tang, Huaqiang Wu, Sylvain Gigan, et al. Training deep physical neural networks with local physical information bottleneck. *arXiv preprint arXiv:2602.09569*, 2026.
- [55] Fabian Böhm, Diego Alonso-Urquijo, Guy Verschaffelt, and Guy Van der Sande. Noise-injected analog Ising machines enable ultrafast statistical sampling and machine learning. *Nature Communications*, 13:5847, 2022.
- [56] Cyrill Bösch, Geoffrey Roeder, Marc Serra-Garcia, and Ryan P Adams. Local learning rules for out-of-equilibrium physical generative models. *arXiv preprint arXiv:2506.19136*, 2025.
- [57] Unifying Physical Backpropagation. <https://github.com/yigithangediz19/Unifying-Physical-Backpropagation->, 2026. GitHub repository.
- [58] Nick Richardson, Cyrill Bösch, and Ryan P Adams. Nonlinear computation with linear optics via source-position encoding. *arXiv preprint arXiv:2504.20401*, 2025.
- [59] Fei Xia, Kyungduk Kim, Yaniv Eliezer, SeungYun Han, Liam Shaughnessy, Sylvain Gigan, and Hui Cao. Nonlinear optical encoding enabled by recurrent linear scattering. *Nature Photonics*, 18(10):1067–1075, 2024.
- [60] Mustafa Yildirim, Niyazi Ulas Dinc, Ilker Oguz, Demetri Psaltis, and Christophe Moser. Nonlinear processing with linear optics. *Nature Photonics*, 18(10):1076–1082, 2024.
- [61] Michael B. Giles and Niles A. Pierce. An introduction to the adjoint approach to design. *Flow, Turbulence and Combustion*, 65(3-4):393–415, December 2000.
- [62] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, November 2006.
- [63] Georgios Veronis, Robert W. Dutton, and Shanhui Fan. Method for sensitivity analysis of photonic crystal devices. *Optics Letters*, 29(19):2288–2290, Oct 2004.
- [64] Jan Werschnik and E. K. U. Gross. Quantum optimal control theory. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 40(18):R175–R211, 2007.
- [65] María Soledad Aronna, Joseph Frédéric Bonnans, and Axel Kröner. Optimal control of PDEs in a complex space setting: Application to the Schrödinger equation. *SIAM Journal on Control and Optimization*, 57(2):1390–1412, 2019.
- [66] D. H. Brandwood. A complex gradient operator and its application in adaptive array theory. *IEEE Proceedings F - Communications, Radar and Signal Processing*, 130(1):11–16, 1983.
- [67] Are Hjørungnes and David Gesbert. Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing*, 55(6):2740–2746, 2007.
- [68] Laurent Sorber, Marc Van Barel, and Lieven De Lathauwer. Unconstrained optimization of real functions in complex variables. *SIAM Journal on Optimization*, 22(3):879–898, 2012.
- [69] Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE Constraints*. Springer Netherlands, 2009.
- [70] Benjamin Scellier, Siddhartha Mishra, Yoshua Bengio, and Yann Ollivier. Agnostic physics-driven deep learning. *arXiv preprint arXiv:2205.15021*, 2022.

- [71] Hyeon Kwon, Arunn Suntharalingam, Laureano Bulus-Rossini, and Tsampikos Kottos. In-situ time-domain physical adjoint optimization of complex wave dynamics. *arXiv preprint arXiv:2608.19456*, 2026.
- [72] Marc Berneman and Daniel Hexner. Equilibrium propagation for dissipative dynamics. *arXiv preprint arXiv:2506.20402*, 2025.
- [73] Marc Berneman and Daniel Hexner. Training overdamped dynamics. *arXiv preprint arXiv:2602.19122*, 2026.
- [74] Jérémie Laydevant, Danijela Marković, and Julie Grollier. Training an Ising machine with equilibrium propagation. *Nature Communications*, 15(1):3671, 2024.
- [75] John Guillaumon, William Tuxbury, Cheng-Zhen Wang, Owen Miller, Zin Lin, and Tsampikos Kottos. In-situ adjoint wave control in reconfigurable non-hermitian nonlinear systems. *arXiv preprint arXiv:2608.13503*, 2026.
- [76] Benjamin Scellier, Anirudh Goyal, Jonathan Binas, Thomas Mesnard, and Yoshua Bengio. Generalization of equilibrium propagation to vector field dynamics. *arXiv preprint arXiv:1808.04873*, 2018.
- [77] Axel Laborieux and Friedemann Zenke. Improving equilibrium propagation without weight symmetry through Jacobian homeostasis. In *International Conference on Learning Representations*, 2024.
- [78] Antonino Emanuele Scurria, Dimitri Vanden Abeele, Bortolo Matteo Mognetti, and Serge Massar. Equilibrium propagation for non-conservative systems. *arXiv preprint arXiv:2602.03670*, 2026.
- [79] Zheming Li, Lucas J Fernández-Alcázar, Zin Lin, and Tsampikos Kottos. In-situ adjoint protocols for nonlinear pt-symmetric self-optimizing machines. *arXiv preprint arXiv:2608.16056*, 2026.
- [80] Mathias Fink. Time reversal of ultrasonic fields. I. Basic principles. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 39(5):555–566, 1992.
- [81] Guillaume Pourcel, Debabrota Basu, Maxence Ernoul, and Aditya Gilra. Lagrangian-based equilibrium propagation: generalisation to arbitrary boundary conditions & equivalence with Hamiltonian echo learning. *arXiv preprint arXiv:2506.06248*, 2025.
- [82] Michael Spivak. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. W. A. Benjamin, 1965.
- [83] Axel Laborieux, Maxence Ernoul, Benjamin Scellier, Yoshua Bengio, Julie Grollier, and Damien Querlioz. Scaling equilibrium propagation to deep convnets by drastically reducing its gradient estimator bias. *Frontiers in Neuroscience*, 15:633674, 2021.
- [84] Serge Massar. Equilibrium propagation for learning in Lagrangian dynamical systems. *Physical Review E*, 112(3):035304, 2025.
- [85] Reinhold Remmert. *Theory of complex functions*, volume 122. Springer Science & Business Media, 1991.
- [86] Kelvin Koor, Yixian Qiu, Leong Chuan Kwek, and Patrick Rebentrost. A short tutorial on Wirtinger calculus with applications in quantum information. *arXiv preprint arXiv:2312.04858*, 2023.
- [87] Farshid Ashtiani, Mohamad Hossein Idjadi, and Kwangwoong Kim. Integrated photonic neural network with on-chip backpropagation training. *Nature*, 651(8107):927–932, 2026.

- [88] Jack Kendall, Ross Pantone, Kalpana Manickavasagam, Yoshua Bengio, and Benjamin Scellier. Training end-to-end analog neural networks with equilibrium propagation. *arXiv preprint arXiv:2006.01981*, 2020.
- [89] Axel Laborieux and Friedemann Zenke. Holomorphic equilibrium propagation computes exact gradients through finite size oscillations. *Advances in Neural Information Processing Systems*, 35:12950–12963, 2022.
- [90] Benjamin Scellier and Yoshua Bengio. Equivalence of equilibrium propagation and recurrent backpropagation. *Neural Computation*, 31(2):312–329, 2019.
- [91] Maxence Ernout, Julie Grollier, Damien Querlioz, Yoshua Bengio, and Benjamin Scellier. Updates of equilibrium prop match gradients of backprop through time in an RNN with static input. *Advances in Neural Information Processing Systems*, 32, 2019.
- [92] Vidyesh Rao Anisetti, Ananth Kandala, Benjamin Scellier, and JM Schwarz. Frequency propagation: Multimechanism learning in nonlinear physical networks. *Neural Computation*, 36(4):596–620, 2024.
- [93] Qingshan Wang, Clara C Wanjura, and Florian Marquardt. Training coupled phase oscillators as a neuromorphic platform using equilibrium propagation. *Neuromorphic Computing and Engineering*, 4(3):034014, 2024.
- [94] Nicola Dal Cin, Florian Marquardt, and Clara C Wanjura. Training nonlinear optical neural networks with scattering backpropagation. *arXiv preprint arXiv:2508.11750*, 2025.
- [95] Tobias Helbig, Tobias Hofmann, Stefan Imhof, Mohamed Abdelghany, Tobias Kiessling, Laurens W Molenkamp, Ching Hua Lee, Alexander Szameit, Martin Greiter, and Ronny Thomale. Generalized bulk–boundary correspondence in non-Hermitian topoelectrical circuits. *Nature Physics*, 16(7):747–750, 2020.
- [96] Sebastian Weidemann, Mark Kremer, Tobias Helbig, Tobias Hofmann, Alexander Stegmaier, Martin Greiter, Ronny Thomale, and Alexander Szameit. Topological funneling of light. *Science*, 368(6488):311–314, 2020.
- [97] Lisa M Nash, Dustin Kleckner, Alismari Read, Vincenzo Vitelli, Ari M Turner, and William T M Irvine. Topological mechanics of gyroscopic metamaterials. *Proceedings of the National Academy of Sciences*, 112(47):14495–14500, 2015.
- [98] Romain Fleury, Dimitrios L Sounas, Caleb F Sieck, Michael R Haberman, and Andrea Alù. Sound isolation and giant linear nonreciprocity in a compact acoustic circulator. *Science*, 343(6170):516–519, 2014.
- [99] Colin Scheibner, Anton Souslov, Debarghya Banerjee, Piotr Surówka, William T M Irvine, and Vincenzo Vitelli. Odd elasticity. *Nature Physics*, 16(4):475–480, 2020.
- [100] Hendrik B G Casimir. On Onsager’s principle of microscopic reversibility. *Reviews of Modern Physics*, 17(2-3):343–350, 1945.
- [101] Kohei Kawabata, Ken Shiozaki, Masahito Ueda, and Masatoshi Sato. Symmetry and topology in non-Hermitian physics. *Physical Review X*, 9(4):041015, 2019.
- [102] Martin Brandenbourger, Xander Locsin, Edan Lerner, and Corentin Coulais. Non-reciprocal robotic metamaterials. *Nature Communications*, 10(1):4608, 2019.
- [103] Ananya Ghatak, Martin Brandenbourger, Jasper van Wezel, and Corentin Coulais. Observation of non-Hermitian topology and its bulk–edge correspondence in an active mechanical metamaterial. *Proceedings of the National Academy of Sciences*, 117(47):29561–29568, 2020.

- [104] Lars Onsager. Reciprocal relations in irreversible processes. I. *Physical Review*, 37(4):405–426, 1931.
- [105] Lars Onsager and Stefan Machlup. Fluctuations and irreversible processes. *Physical Review*, 91(6):1505–1512, 1953.
- [106] Nikita Stroeve and Natalia G Berloff. Programmable k-local ising machines and all-optical kolmogorov-arnold networks on photonic platforms. In *CLEO: Fundamental Science*, pages JTU–215. Optica Publishing Group, 2026.
- [107] Cyrill Bösch, Marc Serra-Garcia, Christian Böhm, and Andreas Fichtner. Adjoint computation of Berry phase gradients. *Journal of Sound and Vibration*, 619:119357, 2025.
- [108] Carl M Bender, Bjorn K Berntson, David Parker, and E Samuel. Observation of PT phase transition in a simple mechanical system. *American Journal of Physics*, 81(3):173–179, 2013.
- [109] Stefano Longhi. Quantum-optical analogies using photonic structures. *Laser & Photonics Reviews*, 3(3):243–261, 2009.
- [110] M. D. Feit and J. A. Fleck, Jr. Light propagation in graded-index optical fibers. *Applied Optics*, 17(24):3990–3998, 1978.
- [111] Sia Ghelichkhan, Angus Gibson, D Rhodri Davies, Stephan C Kramer, and David A Ham. Automatic adjoint-based inversion schemes for geodynamics: reconstructing the evolution of Earth’s mantle in space and time. *Geoscientific Model Development*, 17(13):5057–5086, 2024.
- [112] Patrick E Farrell, David A Ham, Simon W Funke, and Marie E Rognes. Automated derivation of the adjoint of high-level transient finite element programs. *SIAM Journal on Scientific Computing*, 35(4):C369–C393, 2013.

Contents of the appendices

A	Derivation of the adjoint state equation and the gradient formula	37
A.1	State equation	38
A.2	Cost function	38
A.3	Differentiability assumptions	38
A.4	Constrained optimization problem	39
A.5	The Lagrangian	39
A.6	Stationarity with respect to the multipliers: the forward problem	39
A.7	Stationarity turns the partial derivative into the gradient	40
A.8	Stationarity with respect to the state: the adjoint problem	40
A.9	The gradient	41
A.10	Integrating the adjoints forward in time	43
B	Physical adjoint for nonlinear second-order systems: time reversal and the damping obstruction	43
B.1	Time reversal requires removing or flipping the damping	43
B.2	The nudged response on the gain-flipped device	44
B.3	Comparison with the adjoint equation: the damping obstruction	45
B.4	The undamped reciprocal case	45
C	Mass-proportional damping	45
C.1	Exponential weighting interchanges damping and gain	46
C.2	Initial conditions, recovery, and the physical algorithm	46

D	Onsager reciprocity	47
D.1	Second-order real-valued systems	47
D.2	First-order real-valued systems	49
E	Adjoint state equation and gradient for first-order systems	50
E.1	Constrained optimization problem	50
E.2	The Lagrangian	51
E.3	Stationarity with respect to the multipliers: the forward problem	51
E.4	Stationarity turns the partial derivative into the gradient	51
E.5	Stationarity with respect to the state: the adjoint problem	51
E.6	The gradient	52
E.7	Integrating the adjoint forward in time	53
F	Wirtinger derivation of the Schrödinger adjoint	53
F.1	Problem and conventions	53
F.2	Augmented Lagrangian and the Wirtinger adjoint	54
F.3	Forward-time conjugated adjoint	55
F.4	Proof of the nonlinear intertwined-adjoint construction	56
G	Second-order systems: gyroscopic $\mathcal{PT}$ mirrors	57
H	Free-space optics	58
I	Details for stationary problems	60
I.1	The stationary adjoint	60
I.2	Linear systems	61
I.3	Real fields: the static adjoint at a fixed point	62
I.4	Layered photonic networks	63
I.5	Equivalence with scattering-matrix learning rules	64
I.6	Tangent-space adjoint for quantum equilibrium states	65
J	How to test a gradient: the Taylor test	70

A Derivation of the adjoint state equation and the gradient formula

In this appendix we derive the adjoint state equation and the gradient formula through the Lagrangian method of PDE-constrained optimization [61, 69], in the Lagrange-multiplier form that is standard for time-dependent inverse problems [62]. The derivation is organized around a single object, the Lagrangian $\mathcal{L}[\mathbf{u}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{p}]$, viewed as a functional of *independent* arguments: the state trajectory $\mathbf{u}$, the multipliers $\mathbf{b}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, and the parameters $\mathbf{p}$. Every ingredient of the adjoint method is one of its partial derivatives:

- stationarity with respect to the multipliers, $\mathcal{L}_{\mathbf{b}} = 0$, $\mathcal{L}_{\boldsymbol{\lambda}} = 0$, $\mathcal{L}_{\boldsymbol{\mu}} = 0$, recovers the *forward problem* (state equation and initial conditions);
- stationarity with respect to the state trajectory, $\mathcal{L}_{\mathbf{u}} = 0$, defines the *adjoint problem*;
- at a point where all four conditions hold, the remaining partial derivative $\mathcal{L}_{\mathbf{p}}$ equals the total derivative $d\chi/d\mathbf{p}$ of the reduced cost, and evaluating it yields the *gradient* as forward–adjoint overlap integrals.

A.1 State equation

Let V be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Here $V = \mathbb{R}^N$ (a discrete system with N degrees of freedom),

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^N x_n y_n = \mathbf{x}^T \mathbf{y}. \quad (109)$$

Let $\mathbf{u}$ be a path

$$\mathbf{u} \in C^2([0, T], V), \quad (110)$$

so that each $\mathbf{u}(t) \in V$. The state equation is

$$\mathbf{M}(\mathbf{p}_M) \ddot{\mathbf{u}}(t) + \mathbf{D}(\mathbf{p}_D) \dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = 0, \quad t \in [0, T], \quad (111)$$

with parametrized initial conditions $\mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0})$, $\dot{\mathbf{u}}(0) = \mathbf{v}_0(\mathbf{p}_{v_0})$. Here $\mathbf{M}(\mathbf{p}_M), \mathbf{D}(\mathbf{p}_D) : V \rightarrow V$ are linear operators, $\mathbf{F}[\cdot, \mathbf{p}_F, t] : V \rightarrow V$ is a possibly nonlinear operator, and $\mathbf{f} : [0, T] \times \mathbb{R}^{M_f} \rightarrow V$. We assume that $\mathbf{M}(\mathbf{p}_M)$ is invertible for all visited $\mathbf{p}_M$. Each term carries its own parameters $\mathbf{p}_x \in \mathbb{R}^{M_x}$ for $x \in \{M, D, F, f, u_0, v_0\}$, collected into a single vector $\mathbf{p} = [\mathbf{p}_M, \mathbf{p}_D, \mathbf{p}_F, \mathbf{p}_f, \mathbf{p}_{u_0}, \mathbf{p}_{v_0}] \in \mathbb{R}^M$ with $M = \sum_x M_x$.

We use round brackets for explicit parameter dependence, e.g. $\mathbf{M}(\mathbf{p}_M)$, and square brackets for functional dependence, e.g. $\mathbf{F}[\mathbf{u}]$.

A.2 Cost function

Given a path $\mathbf{u} \in C^2([0, T], V)$, we define the cost $\chi : C^2([0, T], V) \rightarrow \mathbb{R}$ by

$$\chi[\mathbf{u}] := \int_0^T \theta[\mathbb{P}\mathbf{u}(t)] dt + \psi[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)], \quad (112)$$

where $\mathbb{P}$ is an orthogonal projector of the state onto output degrees of freedom, i.e. $\mathbb{P} = \mathbb{P}^T = \mathbb{P}^2$, and $\theta : V \rightarrow \mathbb{R}$ and $\psi : V \times V \rightarrow \mathbb{R}$ are possibly nonlinear functionals.

A.3 Differentiability assumptions

The dynamics may be non-autonomous, so $\mathbf{F}$ carries an explicit time argument alongside its state and parameter arguments. We require $\mathbf{F}$ to be C^2 in the latter two and continuous in t , and $\mathbf{f}$ to be C^2 in $\mathbf{p}$. The derivatives below are taken at fixed t , which we suppress in this subsection. Concretely, for each fixed $\mathbf{p}$ the map $\mathbf{u} \mapsto \mathbf{F}[\mathbf{u}, \mathbf{p}]$ has a Fréchet derivative $\mathbf{F}_{\mathbf{u}} : V \rightarrow V$, a bounded linear operator satisfying

$$\mathbf{F}[\mathbf{u} + \varepsilon \mathbf{h}, \mathbf{p}] = \mathbf{F}[\mathbf{u}, \mathbf{p}] + \varepsilon \mathbf{F}_{\mathbf{u}}[\mathbf{u}, \mathbf{p}] \mathbf{h} + \mathcal{O}(\varepsilon^2) \quad \forall \mathbf{h} \in V, \quad (113)$$

where $\mathbf{F}_{\mathbf{u}}[\mathbf{u}, \mathbf{p}] \in \mathbb{R}^{N \times N}$ is the Jacobian with entries $\left(\frac{\partial F_i}{\partial u_j} \right)$. Its adjoint $\mathbf{F}_{\mathbf{u}}^T : V \rightarrow V$ is defined by

$$\langle \mathbf{x}, \mathbf{F}_{\mathbf{u}} \mathbf{h} \rangle = \langle \mathbf{F}_{\mathbf{u}}^T \mathbf{x}, \mathbf{h} \rangle \quad \forall \mathbf{x}, \mathbf{h} \in V. \quad (114)$$

This identity is purely algebraic in V , so no boundary terms arise.

Analogously, for each fixed $\mathbf{u}$ the map $\mathbf{p} \mapsto \mathbf{F}[\mathbf{u}, \mathbf{p}]$ has a derivative $\mathbf{F}_{\mathbf{p}}(\mathbf{u}, \mathbf{p}) : \mathbb{R}^M \rightarrow V$ (the $N \times M$ matrix $\left(\frac{\partial F_i}{\partial p_j} \right)$), and for each fixed t the map $\mathbf{p} \mapsto \mathbf{f}(t, \mathbf{p})$ has a derivative $\mathbf{f}_{\mathbf{p}}(t, \mathbf{p}) : \mathbb{R}^M \rightarrow V$. All these derivatives exist and depend continuously on their arguments. By the Picard–Lindelöf theorem, the state equation (111) then has a unique solution $\mathbf{u}(\cdot; \mathbf{p})$ for each $\mathbf{p}$, and this solution depends smoothly on $\mathbf{p}$.

A.4 Constrained optimization problem

The constrained optimization problem is the following.

Problem A.1 (Constrained optimization of a second-order ODE). Find

$$\min_{\mathbf{p} \in \mathbb{R}^M} \chi[\mathbf{u}] \quad \text{subject to} \quad \begin{cases} \mathbf{M}(\mathbf{p}_M) \ddot{\mathbf{u}}(t) + \mathbf{D}(\mathbf{p}_D) \dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] = \mathbf{f}(t, \mathbf{p}_f), & t \in [0, T], \\ \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}), \\ \dot{\mathbf{u}}(0) = \mathbf{v}_0(\mathbf{p}_{v_0}). \end{cases} \quad (115)$$

The adjoint method computes the gradient of the cost function, $d\chi/d\mathbf{p}$, at the cost of one additional ODE solve, independent of the number of parameters M . We can then solve the optimization problem, for example, by gradient descent:

$$\mathbf{p} \leftarrow \mathbf{p} - \alpha \frac{d\chi}{d\mathbf{p}}, \quad (116)$$

where $\alpha > 0$ is the learning rate.

A.5 The Lagrangian

We enforce the state equation with a multiplier path $\mathbf{b} \in C^2([0, T], V)$ and the two initial conditions with multiplier vectors $\boldsymbol{\lambda}, \boldsymbol{\mu} \in V$, and define

$$\begin{aligned} \mathcal{L}[\mathbf{u}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{p}] := & \int_0^T \langle \mathbf{b}(t), \mathbf{M}(\mathbf{p}_M) \ddot{\mathbf{u}}(t) + \mathbf{D}(\mathbf{p}_D) \dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) \rangle dt \\ & + \langle \boldsymbol{\lambda}, \mathbf{u}(0) - \mathbf{u}_0(\mathbf{p}_{u_0}) \rangle + \langle \boldsymbol{\mu}, \dot{\mathbf{u}}(0) - \mathbf{v}_0(\mathbf{p}_{v_0}) \rangle + \chi[\mathbf{u}]. \end{aligned} \quad (117)$$

This is a functional $\mathcal{L} : C^2([0, T], V) \times C^2([0, T], V) \times V \times V \times \mathbb{R}^M \rightarrow \mathbb{R}$.

The five arguments of $\mathcal{L}$ are *independent* variables: $\mathcal{L}$ is defined on the full, unconstrained product space. In particular, $\mathbf{u}$ is an arbitrary path and is *not* assumed to solve the state equation, while $\mathbf{p}$ enters only through the explicit dependencies $\mathbf{M}(\mathbf{p}_M)$, $\mathbf{D}(\mathbf{p}_D)$, $\mathbf{F}[\cdot, \mathbf{p}_F, \cdot]$, $\mathbf{f}(\cdot, \mathbf{p}_f)$, $\mathbf{u}_0(\mathbf{p}_{u_0})$, $\mathbf{v}_0(\mathbf{p}_{v_0})$. We may therefore differentiate $\mathcal{L}$ with respect to each argument separately, holding the other four fixed. These *partial* (Fréchet) derivatives are denoted $\mathcal{L}_{\mathbf{b}}$, $\mathcal{L}_{\boldsymbol{\lambda}}$, $\mathcal{L}_{\boldsymbol{\mu}}$, $\mathcal{L}_{\mathbf{u}}$, and $\mathcal{L}_{\mathbf{p}}$. We compute these derivatives in turn.

A.6 Stationarity with respect to the multipliers: the forward problem

The Lagrangian is affine in $\mathbf{b}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$, so these three derivatives can be read off directly. Identifying them with elements of the respective spaces through the inner product,

$$\mathcal{L}_{\mathbf{b}}(t) = \mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{D}\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f), \quad (118)$$

and

$$\mathcal{L}_{\boldsymbol{\lambda}} = \mathbf{u}(0) - \mathbf{u}_0(\mathbf{p}_{u_0}), \quad \mathcal{L}_{\boldsymbol{\mu}} = \dot{\mathbf{u}}(0) - \mathbf{v}_0(\mathbf{p}_{v_0}), \quad (119)$$

so $\mathcal{L}_{\mathbf{b}} = 0$ holds if and only if $\mathbf{u}$ satisfies the state equation (111) on $[0, T]$, and $\mathcal{L}_{\boldsymbol{\lambda}} = 0$, $\mathcal{L}_{\boldsymbol{\mu}} = 0$ if and only if $\mathbf{u}$ satisfies the initial conditions.

Stationarity with respect to a multiplier thus recovers the constraint that the multiplier enforces, which is the generic mechanism of Lagrangian relaxation. The three conditions together state that $\mathbf{u} = \mathbf{u}(\cdot; \mathbf{p})$ is the forward solution.

A.7 Stationarity turns the partial derivative into the gradient

Fix $\mathbf{p}$ and evaluate $\mathcal{L}$ at the forward solution $\mathbf{u} = \mathbf{u}(\cdot; \mathbf{p})$. The constraint terms in (117) vanish, so

$$\chi(\mathbf{p}) = \mathcal{L}[\mathbf{u}(\cdot; \mathbf{p}), \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{p}] \quad \forall \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}. \quad (120)$$

Since the multipliers are arbitrary in (120), we may let them, too, depend on $\mathbf{p}$: pick any assignments $\mathbf{b}(\cdot; \mathbf{p})$, $\boldsymbol{\lambda}(\mathbf{p})$, $\boldsymbol{\mu}(\mathbf{p})$ that depend differentiably on $\mathbf{p}$ (the adjoint state constructed below does). Differentiating (120) with respect to $\mathbf{p}$ by the chain rule, one term per argument of $\mathcal{L}$, gives

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{u}} \frac{d\mathbf{u}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{b}} \frac{d\mathbf{b}}{d\mathbf{p}} + \mathcal{L}_{\boldsymbol{\lambda}} \frac{d\boldsymbol{\lambda}}{d\mathbf{p}} + \mathcal{L}_{\boldsymbol{\mu}} \frac{d\boldsymbol{\mu}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{p}}, \quad (121)$$

with every partial derivative evaluated at the point $(\mathbf{u}(\cdot; \mathbf{p}), \mathbf{b}(\cdot; \mathbf{p}), \boldsymbol{\lambda}(\mathbf{p}), \boldsymbol{\mu}(\mathbf{p}), \mathbf{p})$. The left-hand side is the *total* derivative of the reduced cost. The right-hand side contains *partial* derivatives of $\mathcal{L}$ multiplying the sensitivities of its arguments.

Now inspect (121) term by term. The three multiplier terms vanish because $\mathbf{u}(\cdot; \mathbf{p})$ is the forward solution: by Sec. A.6, $\mathcal{L}_{\mathbf{b}} = 0$, $\mathcal{L}_{\boldsymbol{\lambda}} = 0$, $\mathcal{L}_{\boldsymbol{\mu}} = 0$ at this point, whatever the sensitivities $d\mathbf{b}/d\mathbf{p}$, $d\boldsymbol{\lambda}/d\mathbf{p}$, $d\boldsymbol{\mu}/d\mathbf{p}$ may be. The remaining obstruction is the first term. Computing the sensitivity $d\mathbf{u}/d\mathbf{p}$ requires differentiating through the differential equation solver, with one linearized solve per parameter. The M solves are in general prohibitively expensive. But it enters (121) only multiplied by $\mathcal{L}_{\mathbf{u}}$, and the multipliers $\mathbf{b}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ are still free. If we choose them such that

$$\mathcal{L}_{\mathbf{u}} = 0 \quad \text{at } \mathbf{u} = \mathbf{u}(\cdot; \mathbf{p}), \quad (122)$$

then $d\mathbf{u}/d\mathbf{p}$ is never needed, and

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{p}} : \quad (123)$$

the total derivative of the cost equals the partial derivative of the Lagrangian, which involves only the explicit parameter dependence of $\mathbf{M}$, $\mathbf{D}$, $\mathbf{F}$, $\mathbf{f}$, $\mathbf{u}_0$, $\mathbf{v}_0$ and is inexpensive to evaluate. The condition (122) is a linear equation for $(\mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu})$. We now show that it has exactly one solution, the *adjoint state*.

A.8 Stationarity with respect to the state: the adjoint problem

Let $\mathbf{h} \in C^2([0, T], V)$ be an arbitrary test path with no constraints at either endpoint, and compute

$$\mathcal{L}_{\mathbf{u}} \mathbf{h} := \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{L}[\mathbf{u} + \varepsilon \mathbf{h}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{p}] - \mathcal{L}[\mathbf{u}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{p}]}{\varepsilon}. \quad (124)$$

The Lagrangian (117) has three groups of terms. We vary each in turn.

(i) The constraint integral. Replacing $\mathbf{u} \rightarrow \mathbf{u} + \varepsilon \mathbf{h}$ and using the linearity of $\mathbf{M}$, $\mathbf{D}$ and the differentiability (113) of $\mathbf{F}$ gives the following variation of the first term of (117) (the drive $\mathbf{f}$ has no $\mathbf{u}$ -dependence and does not contribute):

$$\int_0^T \langle \mathbf{b}, \mathbf{M}\ddot{\mathbf{h}} + \mathbf{D}\dot{\mathbf{h}} + \mathbf{F}_{\mathbf{u}}\mathbf{h} \rangle dt. \quad (125)$$

To read off conditions on $\mathbf{b}$, all terms must appear as $\langle \cdot, \mathbf{h} \rangle$: we move the operators across the inner product using $\langle \mathbf{x}, \mathbf{M}\mathbf{y} \rangle = \langle \mathbf{M}^T \mathbf{x}, \mathbf{y} \rangle$ (and (114) for $\mathbf{F}_{\mathbf{u}}$), and move the time derivatives off $\mathbf{h}$ and onto $\mathbf{b}$ via integration by parts, twice for the mass term and once for the damping term:

$$\int_0^T \langle \mathbf{M}^T \mathbf{b}, \ddot{\mathbf{h}} \rangle dt = \int_0^T \langle \mathbf{M}^T \ddot{\mathbf{b}}, \mathbf{h} \rangle dt + \left[\langle \mathbf{M}^T \mathbf{b}, \dot{\mathbf{h}} \rangle - \langle \mathbf{M}^T \dot{\mathbf{b}}, \mathbf{h} \rangle \right]_0^T, \quad (126)$$

$$\int_0^T \langle \mathbf{D}^T \mathbf{b}, \dot{\mathbf{h}} \rangle dt = - \int_0^T \langle \mathbf{D}^T \dot{\mathbf{b}}, \mathbf{h} \rangle dt + \left[\langle \mathbf{D}^T \mathbf{b}, \mathbf{h} \rangle \right]_0^T. \quad (127)$$

Combining these terms gives the variation of the constraint integral:

$$\int_0^T \langle \mathbf{M}^T \ddot{\mathbf{b}} - \mathbf{D}^T \dot{\mathbf{b}} + \mathbf{F}_{\mathbf{u}}^T \mathbf{b}, \mathbf{h} \rangle dt + \left[\langle \mathbf{D}^T \mathbf{b} - \mathbf{M}^T \dot{\mathbf{b}}, \mathbf{h} \rangle + \langle \mathbf{M}^T \mathbf{b}, \dot{\mathbf{h}} \rangle \right]_0^T. \quad (128)$$

(ii) **The initial-condition multiplier terms.** Under $\mathbf{u} \rightarrow \mathbf{u} + \varepsilon \mathbf{h}$, the variation is simply

$$\langle \boldsymbol{\lambda}, \mathbf{h}(0) \rangle + \langle \boldsymbol{\mu}, \dot{\mathbf{h}}(0) \rangle. \quad (129)$$

(iii) **The cost.** The differentiability of θ and ψ , together with $\mathbb{P}^T = \mathbb{P}$, gives the cost variation:

$$\int_0^T \langle \mathbb{P} \theta_{\mathbf{u}}, \mathbf{h} \rangle dt + \langle \mathbb{P} \psi_{\mathbf{u}}, \mathbf{h}(T) \rangle + \langle \mathbb{P} \psi_{\dot{\mathbf{u}}}, \dot{\mathbf{h}}(T) \rangle. \quad (130)$$

The full variation. Adding Eqs. (128), (129), (130) and grouping terms by where $\mathbf{h}$ and $\dot{\mathbf{h}}$ are evaluated gives

$$\begin{aligned} \mathcal{L}_{\mathbf{u}} \mathbf{h} = & \int_0^T \langle \mathbf{M}^T \ddot{\mathbf{b}} - \mathbf{D}^T \dot{\mathbf{b}} + \mathbf{F}_{\mathbf{u}}^T \mathbf{b} + \mathbb{P} \theta_{\mathbf{u}}, \mathbf{h} \rangle dt \\ & + \langle \mathbf{D}^T \mathbf{b}(T) - \mathbf{M}^T \dot{\mathbf{b}}(T) + \mathbb{P} \psi_{\mathbf{u}}, \mathbf{h}(T) \rangle + \langle \mathbf{M}^T \mathbf{b}(T) + \mathbb{P} \psi_{\dot{\mathbf{u}}}, \dot{\mathbf{h}}(T) \rangle \\ & + \langle \mathbf{M}^T \dot{\mathbf{b}}(0) - \mathbf{D}^T \mathbf{b}(0) + \boldsymbol{\lambda}, \mathbf{h}(0) \rangle + \langle -\mathbf{M}^T \mathbf{b}(0) + \boldsymbol{\mu}, \dot{\mathbf{h}}(0) \rangle. \end{aligned} \quad (131)$$

Reading off the conditions. Since $\mathbf{h}$ is arbitrary and unconstrained at both endpoints, the values $\mathbf{h}(t)$, $\mathbf{h}(0)$, $\dot{\mathbf{h}}(0)$, $\mathbf{h}(T)$, $\dot{\mathbf{h}}(T)$ can be chosen independently, so $\mathcal{L}_{\mathbf{u}} = 0$ requires each inner product in (131) to vanish separately. The bulk term and the two terms at $t = T$ involve only $\mathbf{b}$. Together they form a terminal-value problem that determines $\mathbf{b}$ uniquely and defines the adjoint state.

Definition A.2 (Adjoint state for the second-order problem). The *adjoint state* $\mathbf{b}(t) \in V$ associated with Problem A.1 is the solution of the adjoint ODE

$$\mathbf{M}^T \ddot{\mathbf{b}}(t) - \mathbf{D}^T \dot{\mathbf{b}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(t), \mathbf{p}_F, t] \mathbf{b}(t) + \mathbb{P} \theta_{\mathbf{u}}[\mathbb{P} \mathbf{u}(t)] = 0, \quad t \in [0, T], \quad (132)$$

a second-order ODE requiring two conditions, with the terminal conditions

$$\mathbf{M}^T \mathbf{b}(T) = -\mathbb{P} \psi_{\dot{\mathbf{u}}}[\mathbb{P} \mathbf{u}(T), \mathbb{P} \dot{\mathbf{u}}(T)], \quad (133)$$

$$\mathbf{M}^T \dot{\mathbf{b}}(T) = \mathbf{D}^T \mathbf{b}(T) + \mathbb{P} \psi_{\mathbf{u}}[\mathbb{P} \mathbf{u}(T), \mathbb{P} \dot{\mathbf{u}}(T)]. \quad (134)$$

We solve the adjoint ODE (132) backward from T to 0 using these terminal conditions. This produces $\mathbf{b}(0)$ and $\dot{\mathbf{b}}(0)$, which in turn determine the multipliers via the two terms at $t = 0$ in (131):

$$\boldsymbol{\lambda} = -\mathbf{M}^T \dot{\mathbf{b}}(0) + \mathbf{D}^T \mathbf{b}(0), \quad (135)$$

$$\boldsymbol{\mu} = \mathbf{M}^T \mathbf{b}(0). \quad (136)$$

The adjoint ODE is linear in $\mathbf{b}$ (the nonlinear forces enter only through their Jacobian, evaluated along the forward trajectory), its coefficients depend continuously on $\mathbf{p}$, and (135)–(136) are explicit. Hence $(\mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is the unique solution of $\mathcal{L}_{\mathbf{u}} = 0$ and depends smoothly on $\mathbf{p}$, as required in Sec. A.7.

A.9 The gradient

The gradient follows by evaluating the remaining partial derivative of $\mathcal{L}$, namely $\mathcal{L}_{\mathbf{p}}$.

Theorem A.3 (Adjoint gradient for the second-order problem). *Let $\mathbf{u} = \mathbf{u}(\cdot; \mathbf{p})$ solve the forward problem of Problem A.1 and let $\mathbf{b}$ be the adjoint state of Definition A.2. Then the gradient of the reduced cost with respect to $\mathbf{p}$ is*

$$\begin{aligned} \frac{d\chi}{d\mathbf{p}} = & \int_0^T \langle \mathbf{b}(t), \mathbf{M}_{\mathbf{p}_M} \ddot{\mathbf{u}}(t) + \mathbf{D}_{\mathbf{p}_D} \dot{\mathbf{u}}(t) + \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}) \rangle dt \\ & + \langle \mathbf{M}^T \dot{\mathbf{b}}(0) - \mathbf{D}^T \mathbf{b}(0), \frac{d\mathbf{u}_0}{d\mathbf{p}_{u_0}} \rangle - \langle \mathbf{M}^T \mathbf{b}(0), \frac{d\mathbf{v}_0}{d\mathbf{p}_{v_0}} \rangle. \end{aligned} \quad (137)$$

Proof. With $(\mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ chosen as in Definition A.2 and Eqs. (135)–(136), we have $\mathcal{L}_{\mathbf{u}} = 0$, and the key identity (123) gives $d\chi/d\mathbf{p} = \mathcal{L}_{\mathbf{p}}$. The explicit $\mathbf{p}$ -dependence of $\mathcal{L}$ enters through $\mathbf{M}(\mathbf{p}_M)$, $\mathbf{D}(\mathbf{p}_D)$, $\mathbf{F}[\mathbf{u}, \mathbf{p}_F, t]$, $\mathbf{f}(t, \mathbf{p}_f)$ in the constraint integral and through $\mathbf{u}_0(\mathbf{p}_{u_0})$, $\mathbf{v}_0(\mathbf{p}_{v_0})$ in the initial-condition terms. The cost χ has no explicit parameter dependence. Differentiating each term at fixed $\mathbf{u}, \mathbf{b}, \boldsymbol{\lambda}, \boldsymbol{\mu}$ gives

$$\mathcal{L}_{\mathbf{p}} = \int_0^T \langle \mathbf{b}(t), \mathbf{M}_{\mathbf{p}_M} \ddot{\mathbf{u}}(t) + \mathbf{D}_{\mathbf{p}_D} \dot{\mathbf{u}}(t) + \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}) \rangle dt - \langle \boldsymbol{\lambda}, \frac{d\mathbf{u}_0}{d\mathbf{p}_{u_0}} \rangle - \langle \boldsymbol{\mu}, \frac{d\mathbf{v}_0}{d\mathbf{p}_{v_0}} \rangle. \quad (138)$$

Substituting the expressions (135)–(136) for $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ gives (137). $\square$

Every quantity on the right-hand side of (137) is known: $\mathbf{u}(t)$ from the forward solve, $\mathbf{b}(t)$ from the backward solve, and the partial derivatives $\mathbf{M}_{\mathbf{p}_M}$, $\mathbf{D}_{\mathbf{p}_D}$, $\mathbf{F}_{\mathbf{p}_F}$, $\mathbf{f}_{\mathbf{p}_f}$, $d\mathbf{u}_0/d\mathbf{p}_{u_0}$, $d\mathbf{v}_0/d\mathbf{p}_{v_0}$ from the model. The cost of evaluating (137) is one forward solve plus one backward solve, independent of the number of parameters M .

If $\mathbf{M}$ and $\mathbf{D}$ do not depend on $\mathbf{p}$ and the initial conditions are $\mathbf{p}$ -independent, Eq. (137) reduces to

$$\frac{d\chi}{d\mathbf{p}} = \int_0^T \langle \mathbf{b}(t), \mathbf{F}_{\mathbf{p}}(\mathbf{u}(t), \mathbf{p}) - \mathbf{f}_{\mathbf{p}}(t, \mathbf{p}) \rangle dt. \quad (139)$$

Interpreting $\langle \mathbf{b}, \mathbf{F}_{\mathbf{p}} \rangle$. Recall that $\mathbf{F}_{\mathbf{p}} : \mathbb{R}^M \rightarrow V$ is a linear *operator*, not an element of V , so the symbol $\langle \mathbf{b}, \mathbf{F}_{\mathbf{p}} \rangle$ is a shorthand for an M -component vector, to be read componentwise: for each parameter p_j , the partial derivative $\partial \mathbf{F} / \partial p_j \in V$ is a vector, and the j -th component of the gradient is the scalar inner product

$$\frac{d\chi}{dp_j} = \int_0^T \left\langle \mathbf{b}(t), \frac{\partial \mathbf{F}}{\partial p_j}[\mathbf{u}(t), \mathbf{p}] - \frac{\partial \mathbf{f}}{\partial p_j}(t, \mathbf{p}) \right\rangle dt. \quad (140)$$

Equivalently, using the adjoint operator $\mathbf{F}_{\mathbf{p}}^T : V \rightarrow \mathbb{R}^M$ defined by $\langle \mathbf{b}, \mathbf{F}_{\mathbf{p}} \mathbf{q} \rangle_V = \langle \mathbf{F}_{\mathbf{p}}^T \mathbf{b}, \mathbf{q} \rangle_{\mathbb{R}^M}$ for all $\mathbf{q} \in \mathbb{R}^M$, the compact form is

$$\frac{d\chi}{d\mathbf{p}} = \int_0^T (\mathbf{F}_{\mathbf{p}}^T \mathbf{b}(t) - \mathbf{f}_{\mathbf{p}}^T \mathbf{b}(t)) dt \in \mathbb{R}^M. \quad (141)$$

Hence $\mathbf{F}_{\mathbf{p}}^T \mathbf{b}$ is the M -vector obtained by multiplying the transpose of the $N \times M$ Jacobian by the N -vector $\mathbf{b}$. The same interpretation applies to the $\mathbf{F}_{\mathbf{p}_F}$, $\mathbf{M}_{\mathbf{p}_M}$, $\mathbf{D}_{\mathbf{p}_D}$, and $\mathbf{f}_{\mathbf{p}_f}$ terms in the full gradient (137).

Direct optimization of initial conditions. Suppose a subvector of $\mathbf{p}$ encodes the initial state itself, i.e. $\mathbf{u}_0(\mathbf{p}_{u_0}) = \mathbf{p}_{u_0}$ with $\mathbf{p}_{u_0} \in V$. Then $d\mathbf{u}_0/d\mathbf{p}_{u_0} = I_V$ and $d\mathbf{v}_0/d\mathbf{p}_{u_0} = 0$, so the $\mathbf{u}_0$ -block of Eq. (137) identifies the gradient directly:

$$\frac{d\chi}{d\mathbf{u}_0} = \mathbf{M}^T \dot{\mathbf{b}}(0) - \mathbf{D}^T \mathbf{b}(0). \quad (142)$$

Analogously, if a subvector of $\mathbf{p}$ equals $\mathbf{v}_0$, then $d\mathbf{v}_0/d\mathbf{p}_{v_0} = I_V$, $d\mathbf{u}_0/d\mathbf{p}_{v_0} = 0$, and

$$\frac{d\chi}{d\mathbf{v}_0} = -\mathbf{M}^T \mathbf{b}(0). \quad (143)$$

A.10 Integrating the adjoints forward in time

The adjoint problem of Definition A.2 is posed *backward* in time: the data (133)–(134) are imposed at the final time and the ODE (132) is integrated from T down to 0. Physical hardware, by contrast, runs forward. The change of variable $t \mapsto T - t$ removes the mismatch.

Define the forward-time adjoint field

$$\mathbf{a}(t) := \mathbf{b}(T - t), \quad \dot{\mathbf{a}}(t) = -\dot{\mathbf{b}}(T - t), \quad \ddot{\mathbf{a}}(t) = \ddot{\mathbf{b}}(T - t), \quad (144)$$

so that odd derivatives flip sign under the reversal and even ones do not. Evaluating the adjoint ODE (132) at time $T - t$ and substituting (144) gives the forward-time adjoint equation

$$\mathbf{M}^T \ddot{\mathbf{a}}(t) + \mathbf{D}^T \dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T - t), \mathbf{p}_F, T - t] \mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T - t)] = 0, \quad t \in [0, T]. \quad (145)$$

Among the operators only the damping changes, $-\mathbf{D}^T \rightarrow +\mathbf{D}^T$, with mass and stiffness untouched. This sign flip is the entire algebraic effect of time reversal on the dissipative term, and it is the origin of the obstruction to a physical realization in dissipative systems studied in Appendix B. The linearization and the source are now read along the reversed trajectory $\mathbf{u}(T - t)$.

The terminal data are imposed at the reversed time $t = 0$. With $\mathbf{b}(T) = \mathbf{a}(0)$ and $\dot{\mathbf{b}}(T) = -\dot{\mathbf{a}}(0)$, the terminal conditions (133)–(134) become initial conditions,

$$\mathbf{M}^T \mathbf{a}(0) = -\mathbb{P}\psi_{\dot{\mathbf{u}}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)], \quad (146)$$

$$\mathbf{M}^T \dot{\mathbf{a}}(0) = -\mathbf{D}^T \mathbf{a}(0) - \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)]. \quad (147)$$

Likewise $\mathbf{b}(0) = \mathbf{a}(T)$ and $\dot{\mathbf{b}}(0) = -\dot{\mathbf{a}}(T)$, so substituting $\mathbf{b}(t) = \mathbf{a}(T - t)$ throughout the gradient (137) gives

$$\begin{aligned} \frac{d\chi}{d\mathbf{p}} = \int_0^T \left\langle \mathbf{a}(T - t), \mathbf{M}_{\mathbf{p}_M} \ddot{\mathbf{u}}(t) + \mathbf{D}_{\mathbf{p}_D} \dot{\mathbf{u}}(t) + \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}) \right\rangle dt \\ - \left\langle \mathbf{M}^T \dot{\mathbf{a}}(T) + \mathbf{D}^T \mathbf{a}(T), \frac{d\mathbf{u}_0}{d\mathbf{p}_{u_0}} \right\rangle - \left\langle \mathbf{M}^T \mathbf{a}(T), \frac{d\mathbf{v}_0}{d\mathbf{p}_{v_0}} \right\rangle. \end{aligned} \quad (148)$$

Passing to forward time therefore does three things: it flips the sign of the damping operator, it reverses the trajectory argument $\mathbf{u}(t) \mapsto \mathbf{u}(T - t)$ feeding the linearization and the source, and it turns the terminal data at $t = T$ into initial data at $t = 0$.

Eqs. (145)–(148) are the adjoint problem and gradient formula quoted in the main text.

B Physical adjoint for nonlinear second-order systems: time reversal and the damping obstruction

This appendix derives the nudged time-reversed dynamics quoted in Sec. 5, in particular the gain-flipped perturbation equation (28), and proves the nonlinear part of Theorem 5.1. Throughout we use the forward-time adjoint equation (4), which propagates with a damping term $+\mathbf{D}^T \dot{\mathbf{a}}$, an internal-force linearization $\mathbf{F}_{\mathbf{u}}^T$ evaluated along the time-reversed trajectory $\mathbf{u}(T - t)$, the source (5), and initial conditions (6)–(7) set by the terminal-loss derivatives $\mathbb{P}\psi_{\mathbf{u}}$ and $\mathbb{P}\psi_{\dot{\mathbf{u}}}$.

B.1 Time reversal requires removing or flipping the damping

Unlike the linear case, in the nonlinear adjoint (4) the field $\mathbf{a}$ propagates under $\mathbf{F}_{\mathbf{u}}^T$ evaluated along the *reversed* trajectory $\mathbf{u}(T - t)$. To access this linearization physically, we must first reproduce the reversed trajectory on the device. Define

$$\mathbf{w}(t) := \mathbf{u}(T - t), \quad \dot{\mathbf{w}}(t) = -\dot{\mathbf{u}}(T - t), \quad \ddot{\mathbf{w}}(t) = \ddot{\mathbf{u}}(T - t). \quad (149)$$

Evaluating the forward state equation (2) at $T - t$ and using (149) gives the reversed state equation (24),

$$\mathbf{M}\ddot{\mathbf{w}}(t) - \mathbf{D}\dot{\mathbf{w}}(t) + \mathbf{F}[\mathbf{w}(t), \mathbf{p}_F, T - t] - \mathbf{f}(T - t, \mathbf{p}_f) = \mathbf{0}, \quad (150)$$

with initial conditions

$$\mathbf{w}(0) = \mathbf{u}(T), \quad \dot{\mathbf{w}}(0) = -\dot{\mathbf{u}}(T). \quad (151)$$

The only difference from the forward equation (2) is the sign of the damping term: time reversal sends $\mathbf{D}\dot{\mathbf{u}} \mapsto -\mathbf{D}\dot{\mathbf{w}}$, because the first derivative is odd under $t \mapsto T - t$ while the second derivative is even. Reproducing $\mathbf{w}$ on the *same* hardware therefore requires either $\mathbf{D} = \mathbf{0}$ or an experimental replacement $\mathbf{D} \mapsto -\mathbf{D}$ that turns damping into gain. We grant the latter possibility in the next subsection and show that it still fails to deliver the adjoint.

B.2 The nudged response on the gain-flipped device

Suppose the gain flip $\mathbf{D} \mapsto -\mathbf{D}$ can be realized, so that (150) runs on the device and produces the reversed background $\mathbf{w}$. On this same gain-flipped device we superimpose an infinitesimal nudge: a source in the direction of the adjoint source (5) and a perturbation of the initial conditions in the directions of the adjoint terminal data (6)–(7). The nudged field $\mathbf{w}^\epsilon$ obeys

$$\mathbf{M}\ddot{\mathbf{w}}^\epsilon(t) - \mathbf{D}\dot{\mathbf{w}}^\epsilon(t) + \mathbf{F}[\mathbf{w}^\epsilon(t), \mathbf{p}_F, T - t] - \mathbf{f}(T - t, \mathbf{p}_f) + \epsilon \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T - t)] = \mathbf{0}, \quad (152)$$

with initial conditions

$$\mathbf{w}^\epsilon(0) = \mathbf{u}(T) + \epsilon \mathbf{a}(0), \quad \dot{\mathbf{w}}^\epsilon(0) = -\dot{\mathbf{u}}(T) + \epsilon \dot{\mathbf{a}}(0), \quad |\epsilon| \ll 1, \quad (153)$$

where $\mathbf{a}(0)$ and $\dot{\mathbf{a}}(0)$ are obtained by solving the adjoint initial conditions (6)–(7):

$$\mathbf{a}(0) = -\mathbf{M}^{-1}\mathbb{P}\psi_{\dot{\mathbf{u}}}, \quad (154)$$

$$\dot{\mathbf{a}}(0) = -\mathbf{M}^{-1}[\mathbf{D}^T \mathbf{a}(0) + \mathbb{P}\psi_{\mathbf{u}}] = \mathbf{M}^{-1}[\mathbf{D}^T \mathbf{M}^{-1}\mathbb{P}\psi_{\dot{\mathbf{u}}} - \mathbb{P}\psi_{\mathbf{u}}], \quad (155)$$

where we set $\mathbf{M}^T = \mathbf{M}$. The velocity nudge carries the damping contribution $\mathbf{D}^T \mathbf{a}(0)$ demanded by (7), which drops out when $\mathbf{D} = \mathbf{0}$. Write the differential response

$$\delta \mathbf{w}(t) := \mathbf{w}^\epsilon(t) - \mathbf{w}(t) = \mathcal{O}(\epsilon). \quad (156)$$

Expanding the internal force to first order,

$$\mathbf{F}[\mathbf{w}^\epsilon, \mathbf{p}_F, T - t] = \mathbf{F}[\mathbf{w}, \mathbf{p}_F, T - t] + \mathbf{F}_{\mathbf{u}}[\mathbf{w}(t), \mathbf{p}_F, T - t] \delta \mathbf{w}(t) + \mathcal{O}(\|\delta \mathbf{w}\|^2), \quad (157)$$

and subtracting the reversed state equation (150) from (152) gives the gain-flipped perturbation equation (28),

$$\mathbf{M}\delta \ddot{\mathbf{w}}(t) - \mathbf{D}\delta \dot{\mathbf{w}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{w}(t), \mathbf{p}_F, T - t] \delta \mathbf{w}(t) + \epsilon \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T - t)] + \mathcal{O}(\epsilon^2) = \mathbf{0}, \quad (158)$$

with initial conditions obtained by subtracting (151) from (153),

$$\delta \mathbf{w}(0) = \epsilon \mathbf{a}(0), \quad \delta \dot{\mathbf{w}}(0) = \epsilon \dot{\mathbf{a}}(0). \quad (159)$$

Because $\mathbf{w}(t) = \mathbf{u}(T - t)$, the linearization $\mathbf{F}_{\mathbf{u}}[\mathbf{w}(t), \mathbf{p}_F, T - t]$ in (158) is precisely the time-reversed Jacobian $\mathbf{F}_{\mathbf{u}}[\mathbf{u}(T - t), \mathbf{p}_F, T - t]$ appearing in the adjoint equation (4).

B.3 Comparison with the adjoint equation: the damping obstruction

Introduce the rescaled response $\tilde{\mathbf{a}}(t) := \delta \mathbf{w}(t)/\epsilon$, which is $\mathcal{O}(1)$. Dividing (158)–(159) by ϵ and letting $\epsilon \rightarrow 0$ gives

$$\mathbf{M}\ddot{\tilde{\mathbf{a}}}(t) - \mathbf{D}\dot{\tilde{\mathbf{a}}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\tilde{\mathbf{a}}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}, \quad (160)$$

with $\tilde{\mathbf{a}}(0) = \mathbf{a}(0)$ and $\dot{\tilde{\mathbf{a}}}(0) = \dot{\mathbf{a}}(0)$, given by (154)–(155). We compare this term by term with the forward-time adjoint equation (4),

$$\mathbf{M}^T\ddot{\mathbf{a}}(t) + \mathbf{D}^T\dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}, \quad (161)$$

whose initial conditions (6)–(7) the nudge (153) reproduces by construction. Matching the propagators in (160) and (161) requires:

- **Mass.** $\mathbf{M}^T = \mathbf{M}$, so the inertial operators agree.
- **Stiffness.** $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ along the reversed trajectory, so the two internal-force linearizations agree, i.e. the tangent dynamics is reciprocal.
- **Damping.** The rescaled response carries $-\mathbf{D}\dot{\tilde{\mathbf{a}}}$, whereas the adjoint carries $+\mathbf{D}^T\dot{\mathbf{a}}$. For reciprocal damping $\mathbf{D}^T = \mathbf{D}$ the two terms differ by $2\mathbf{D}\dot{\mathbf{a}}$, and they coincide *only* when $\mathbf{D} = \mathbf{0}$.

Even after granting the gain flip needed to reproduce the reversed background, the damping enters the nudged perturbation with the sign $-\mathbf{D}$, the opposite of the $+\mathbf{D}^T$ demanded by the adjoint. A discrepancy in the propagator cannot be repaired by any choice of source or initial condition, so the same hardware cannot simultaneously carry the reversed background and propagate the correct adjoint perturbation around it unless the damping vanishes.

B.4 The undamped reciprocal case

Set $\mathbf{D} = \mathbf{0}$ and assume $\mathbf{M}^T = \mathbf{M}$ together with $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$ along the trajectory. Then the reversed background (150) runs on the same hardware, and the rescaled perturbation equation (160) coincides with the adjoint equation (161), whose initial conditions the nudge already carries. Here (154)–(155) reduce to $\mathbf{a}(0) = -\mathbf{M}^{-1}\mathbb{P}\psi_{\mathbf{u}}$ and $\dot{\mathbf{a}}(0) = -\mathbf{M}^{-1}\mathbb{P}\psi_{\mathbf{u}}$, the nudge quoted in the main text. By uniqueness of solutions, $\tilde{\mathbf{a}} = \mathbf{a}$, i.e.

$$\mathbf{a}(t) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(t) - \mathbf{w}(t)}{\epsilon}, \quad \mathbf{w}(t) = \mathbf{u}(T-t). \quad (162)$$

The adjoint field is therefore the infinitesimal differential response of the nudged time-reversed trajectory, and the gradient follows from the overlap integrals (8)–(13). This proves the nonlinear part of Theorem 5.1. $\square$

C Mass-proportional damping

The obstruction established in Appendix B forbids damping in the nonlinear case because the nudged response on the gain-flipped device propagates with $-\mathbf{D}\dot{\tilde{\mathbf{a}}}$, while the adjoint requires $+\mathbf{D}^T\dot{\mathbf{a}}$. The special case of *mass-proportional* damping reconciles the two. The construction still flips damping to gain, so the adjoint propagates on a device other than the one that ran the forward pass, which places it outside the *same-device* constructions of this article. The mechanism is instructive nonetheless.

Mass-proportional damping means

$$\mathbf{D} = \gamma \mathbf{M}, \quad \gamma \in \mathbb{R} \text{ a known scalar.} \quad (163)$$

Because the damping operator is now a scalar multiple of the inertia, a single scalar exponential weighting in time interchanges the damped and anti-damped equations while leaving $\mathbf{M}$ and the stiffness untouched. We assume, as in Appendix B, reciprocity along the reversed trajectory, $\mathbf{M}^T = \mathbf{M}$ and $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$. Note that (163) makes $\mathbf{D}$ automatically reciprocal, $\mathbf{D}^T = \gamma \mathbf{M}^T = \mathbf{D}$.

C.1 Exponential weighting interchanges damping and gain

With $\mathbf{D} = \gamma\mathbf{M}$ the forward-time adjoint equation (4) reads

$$\mathbf{M}\ddot{\mathbf{a}}(t) + \gamma\mathbf{M}\dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}, \quad (164)$$

i.e. a *damped* linear oscillator for $\mathbf{a}$. The gain-flipped device of Appendix B, by contrast, can only realize the *anti-damped* perturbation dynamics

$$\mathbf{M}\ddot{\tilde{\mathbf{a}}}(t) - \gamma\mathbf{M}\dot{\tilde{\mathbf{a}}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\tilde{\mathbf{a}}(t) + \boldsymbol{\sigma}(t) = \mathbf{0}, \quad (165)$$

where $\boldsymbol{\sigma}$ is the source we are free to apply. Substitute the exponential ansatz

$$\mathbf{a}(t) = e^{-\gamma t} \tilde{\mathbf{a}}(t) \quad (166)$$

into (164), using $\dot{\mathbf{a}} = e^{-\gamma t}(\dot{\tilde{\mathbf{a}}} - \gamma\tilde{\mathbf{a}})$ and $\ddot{\mathbf{a}} = e^{-\gamma t}(\ddot{\tilde{\mathbf{a}}} - 2\gamma\dot{\tilde{\mathbf{a}}} + \gamma^2\tilde{\mathbf{a}})$. The $\gamma^2\mathbf{M}$ contributions cancel and the coefficient of $\dot{\tilde{\mathbf{a}}}$ flips from $+\gamma$ to $-\gamma$, leaving

$$e^{-\gamma t}[\mathbf{M}\ddot{\tilde{\mathbf{a}}} - \gamma\mathbf{M}\dot{\tilde{\mathbf{a}}} + \mathbf{F}_{\mathbf{u}}\tilde{\mathbf{a}}] + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}. \quad (167)$$

Multiplying by $e^{\gamma t}$ shows that $\tilde{\mathbf{a}}$ obeys the anti-damped equation (165), provided the applied source is the exponentially weighted adjoint source

$$\boldsymbol{\sigma}(t) = e^{\gamma t} \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)]. \quad (168)$$

The mass and stiffness operators in (165) are identical to those of the device. The gain flip reverses the sign of the damping, and the source carries the known scalar weight $e^{\gamma t}$. For an objective with only a terminal loss ($\theta \equiv 0$), the source vanishes and no weighting of the drive is needed.

C.2 Initial conditions, recovery, and the physical algorithm

The ansatz (166) also maps the adjoint initial conditions (6)–(7) (with $\mathbf{D} = \gamma\mathbf{M}$) onto conditions of the undamped form for $\tilde{\mathbf{a}}$. From $\mathbf{a}(0) = \tilde{\mathbf{a}}(0)$ and $\dot{\mathbf{a}}(0) = \dot{\tilde{\mathbf{a}}}(0) - \gamma\tilde{\mathbf{a}}(0)$, the terms proportional to γ cancel and

$$\mathbf{M}\tilde{\mathbf{a}}(0) = -\mathbb{P}\psi_{\dot{\mathbf{u}}}, \quad \mathbf{M}\dot{\tilde{\mathbf{a}}}(0) = -\mathbb{P}\psi_{\mathbf{u}}, \quad (169)$$

i.e. the same nudge directions as in the undamped case of Appendix B.4. The physically measured field $\tilde{\mathbf{a}}$ is therefore obtained as in the undamped nonlinear construction, with the single modification that the adjoint source is applied with the weight $e^{\gamma t}$. The true adjoint is recovered by undoing the weighting,

$$\mathbf{a}(t) = e^{-\gamma t} \tilde{\mathbf{a}}(t) = e^{-\gamma t} \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(t) - \mathbf{w}(t)}{\epsilon}, \quad \mathbf{w}(t) = \mathbf{u}(T-t), \quad (170)$$

before substitution into the gradient overlaps (8)–(13). Equivalently, the weight can be carried directly into the overlaps: since $\mathbf{a}(T-t) = e^{-\gamma(T-t)}\tilde{\mathbf{a}}(T-t)$, each integrand acquires the known scalar factor $e^{-\gamma(T-t)}$, e.g.

$$\frac{d\chi}{d\mathbf{p}_F} = \int_0^T e^{-\gamma(T-t)} \langle \tilde{\mathbf{a}}(T-t), \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] \rangle dt, \quad (171)$$

which realizes the exponential time weighting referenced in Sec. 5. Concretely:

1. Run the forward dynamics and record $\mathbf{u}(t)$, then form $\mathbf{w}(t) = \mathbf{u}(T-t)$.
2. Flip the damping to gain, $\mathbf{D} = \gamma\mathbf{M} \mapsto -\gamma\mathbf{M}$, so the device reproduces the reversed background $\mathbf{w}$.
3. On the same gain-flipped device, apply the TRM and nudge together: initialize with $\mathbf{w}^\epsilon(0) = \mathbf{u}(T) - \epsilon\mathbf{M}^{-1}\mathbb{P}\psi_{\dot{\mathbf{u}}}$ and $\dot{\mathbf{w}}^\epsilon(0) = -\dot{\mathbf{u}}(T) - \epsilon\mathbf{M}^{-1}\mathbb{P}\psi_{\mathbf{u}}$, drive with $\mathbf{f}(T-t) - \epsilon e^{\gamma t}\mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)]$, and reverse the explicit time dependence of $\mathbf{F}$.

4. Measure $\tilde{\mathbf{a}}(t) = \lim_{\epsilon \rightarrow 0} [\mathbf{w}^\epsilon(t) - \mathbf{w}(t)]/\epsilon$ and form the adjoint $\mathbf{a}(t) = e^{-\gamma t} \tilde{\mathbf{a}}(t)$.
5. Evaluate the gradient from (8)–(13), equivalently (171).

Two remarks. First, the construction requires γ to be a known *scalar*: only then is $e^{-\gamma t}$ a scalar weight that commutes with $\mathbf{M}$ and $\mathbf{F}_{\mathbf{u}}$ and leaves the stiffness untouched in (167). A damping operator $\mathbf{D}$ that is not a scalar multiple of $\mathbf{M}$ would require a matrix exponential $e^{-\mathbf{M}^{-1}\mathbf{D}t}$ that does not, in general, commute with $\mathbf{F}_{\mathbf{u}}$, and the obstruction of Appendix B returns. Second, as in the undamped case, this still requires the gain flip to reproduce the reversed background. Mass-proportional damping only makes the subsequent adjoint extraction exact, through the known exponential reweighting [81].

D Onsager reciprocity

Section 9 identifies Onsager reciprocity as a special case of the general intertwining construction. Here we derive its second- and first-order real-valued forms. The physical field is the weighted adjoint $\mathbf{d} = \mathbf{V}\mathbf{a}$. Whether the gradient requires explicit recovery of $\mathbf{a} = \mathbf{V}^{-1}\mathbf{d}$ depends on how the trainable parameters enter the dynamics.

D.1 Second-order real-valued systems

D.1.1 Assumptions

Consider the second-order system of Sec. 2, with constant mass and damping operators. Let $\mathbf{V} = \mathbf{V}^T \succ 0$ be a constant Onsager mobility. We assume that all three forward operators are $\mathbf{V}$ -reciprocal,

$$\mathbf{V}\mathbf{M}^T\mathbf{V}^{-1} = \mathbf{M}, \quad \mathbf{V}\mathbf{D}^T\mathbf{V}^{-1} = \mathbf{D}, \quad \mathbf{V}\mathbf{F}_{\mathbf{u}}^T\mathbf{V}^{-1} = \mathbf{F}_{\mathbf{u}}, \quad (172)$$

with the last condition evaluated along the (time-reversed) forward trajectory. In the common symmetric case $\mathbf{M}^T = \mathbf{M}$ and $\mathbf{D}^T = \mathbf{D}$, the first two conditions are the commutation relations $\mathbf{M}\mathbf{V} = \mathbf{V}\mathbf{M}$ and $\mathbf{D}\mathbf{V} = \mathbf{V}\mathbf{D}$.

The condition on $\mathbf{F}_{\mathbf{u}}$ is the linearized form of an *Onsager-symmetric gradient flow*: if the internal force derives from a scalar energy through the mobility,

$$\mathbf{F}(\mathbf{u}, \mathbf{p}_F) = \mathbf{V} E_{\mathbf{u}}(\mathbf{u}, \mathbf{p}_F), \quad (173)$$

then $\mathbf{F}_{\mathbf{u}} = \mathbf{V}E_{\mathbf{uu}}$ with the Hessian $E_{\mathbf{uu}}$ symmetric, so $\mathbf{V}\mathbf{F}_{\mathbf{u}}^T\mathbf{V}^{-1} = \mathbf{V}E_{\mathbf{uu}}^T\mathbf{V}^T\mathbf{V}^{-1} = \mathbf{V}E_{\mathbf{uu}} = \mathbf{F}_{\mathbf{u}}$, as required. For a quadratic energy $E = \frac{1}{2}\langle \mathbf{u}, \mathbf{H}\mathbf{u} \rangle$ with $\mathbf{H}^T = \mathbf{H}$ this is the linear Onsager-reciprocal case $\mathbf{F} = \mathbf{V}\mathbf{H}\mathbf{u}$. For nonquadratic E the dynamics is nonlinear, but the mobility is still constant and reciprocal.

D.1.2 The weighted adjoint

Recall the forward-time adjoint equation (4),

$$\mathbf{M}^T\ddot{\mathbf{a}}(t) + \mathbf{D}^T\dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}. \quad (174)$$

Define the transformed field

$$\mathbf{d}(t) := \mathbf{V}\mathbf{a}(t), \quad \text{equivalently} \quad \mathbf{a}(t) = \mathbf{V}^{-1}\mathbf{d}(t). \quad (175)$$

Substituting $\mathbf{a} = \mathbf{V}^{-1}\mathbf{d}$ into (174) and multiplying on the left by $\mathbf{V}$ gives

$$(\mathbf{V}\mathbf{M}^T\mathbf{V}^{-1})\ddot{\mathbf{d}} + (\mathbf{V}\mathbf{D}^T\mathbf{V}^{-1})\dot{\mathbf{d}} + (\mathbf{V}\mathbf{F}_{\mathbf{u}}^T\mathbf{V}^{-1})\mathbf{d} + \mathbf{V}\mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}. \quad (176)$$

By the $\mathbf{V}$ -reciprocity assumptions (172), the three conjugated operators collapse to the *forward* operators $\mathbf{M}$, $\mathbf{D}$, $\mathbf{F}_{\mathbf{u}}$, leaving

$$\mathbf{M}\ddot{\mathbf{d}}(t) + \mathbf{D}\dot{\mathbf{d}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\mathbf{d}(t) + \mathbf{V}\mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}. \quad (177)$$

This is governed by the same linearized operator that the physical device realizes around the (time-reversed) forward trajectory, as in the Euclidean case $\mathbf{V} = \mathbf{I}$, but with the adjoint source now carrying the weight $\mathbf{V}$.

The adjoint initial conditions (6)–(7) transform the same way. Substituting $\mathbf{a} = \mathbf{V}^{-1}\mathbf{d}$ and multiplying by $\mathbf{V}$ gives

$$\mathbf{M}\mathbf{d}(0) = -\mathbf{V}\mathbb{P}\psi_{\dot{\mathbf{u}}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)], \quad (178)$$

$$\mathbf{M}\dot{\mathbf{d}}(0) = -\mathbf{D}\mathbf{d}(0) - \mathbf{V}\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\dot{\mathbf{u}}(T)], \quad (179)$$

again using (172). Thus $\mathbf{d}$ solves the forward-operator problem (177)–(179), with every adjoint source and initial-condition kick simply pre-multiplied by $\mathbf{V}$.

D.1.3 Physical realization and gradient recovery

Eq. (177) has the structure analyzed for $\mathbf{V} = \mathbf{I}$: it propagates under the forward $\mathbf{M}$, $\mathbf{D}$, $\mathbf{F}_{\mathbf{u}}$. The physical constructions therefore apply unchanged to $\mathbf{d}$:

- **Linear $\mathbf{V}$ -reciprocal** ($\mathbf{F} = \mathbf{V}\mathbf{H}\mathbf{u}$, with $\mathbf{V}$ -reciprocal $\mathbf{M}$ and $\mathbf{D}$): $\mathbf{d}$ is obtained by a direct finite-amplitude run of the same hardware, driven by the $\mathbf{V}$ -weighted adjoint source and initialized by (178)–(179). As in the Euclidean case, $\mathbf{V}$ -reciprocal damping is permitted.
- **Nonlinear gradient flow** ($\mathbf{F} = \mathbf{V}E_{\mathbf{u}}$, $\mathbf{D} = \mathbf{0}$): $\mathbf{d}$ is obtained as the infinitesimal response of the nudged time-reversed trajectory (Appendix B), now with the source and initial-condition nudges pre-multiplied by $\mathbf{V}$. If instead $\mathbf{D} = \gamma\mathbf{M}$, the exponential reweighting of Appendix C also applies.

In both cases the measured physical field is $\mathbf{d} = \mathbf{V}\mathbf{a}$. The Euclidean adjoint is

$$\mathbf{a}(t) = \mathbf{V}^{-1}\mathbf{d}(t), \quad (180)$$

so the force-parameter gradient can be written directly in terms of the measured field as

$$\frac{d\chi}{d\mathbf{p}_F} = \int_0^T \langle \mathbf{V}^{-1}\mathbf{d}(T-t), \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] \rangle dt. \quad (181)$$

If $\mathbf{F} = \mathbf{V}E_{\mathbf{u}}$ and $\mathbf{V}$ is independent of the parameters $\mathbf{p}_F$ being differentiated, then $\mathbf{F}_{\mathbf{p}_F} = \mathbf{V}E_{\mathbf{u}\mathbf{p}_F}$. Since $\mathbf{V}$ is symmetric, the weighting cancels:

$$\begin{aligned} \frac{d\chi}{d\mathbf{p}_F} &= \int_0^T \langle \mathbf{V}^{-1}\mathbf{d}(T-t), \mathbf{V}E_{\mathbf{u}\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] \rangle dt \\ &= \int_0^T \langle \mathbf{d}(T-t), E_{\mathbf{u}\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] \rangle dt. \end{aligned} \quad (182)$$

These gradients therefore do not require explicit recovery of $\mathbf{a}$. The same cancellation applies to source parameters when $\mathbf{f} = \mathbf{V}\mathbf{j}$ with fixed $\mathbf{V}$, and to mass or damping parameters when their sensitivities carry the same left factor $\mathbf{V}$. Other parameter sensitivities and initial-condition gradients retain the weights in the general formulas. If the differentiated parameter changes $\mathbf{V}$ itself, its derivative must also be included. This distinction depends on the parameterization, not on the order of the dynamics.

This proves the second-order Onsager-reciprocal extension stated in Sec. 9 and generalizes Theorem 5.1.

D.2 First-order real-valued systems

For the first-order system $\dot{\mathbf{u}} + \mathbf{F}[\mathbf{u}, \mathbf{p}_F, t] - \mathbf{f} = \mathbf{0}$, assume $\mathbf{V}\mathbf{F}_{\mathbf{u}}^T\mathbf{V}^{-1} = \mathbf{F}_{\mathbf{u}}$ with constant $\mathbf{V} = \mathbf{V}^T \succ 0$. Substituting $\mathbf{d} = \mathbf{V}\mathbf{a}$ into the first-order adjoint (211) and its initial condition gives

$$\dot{\mathbf{d}}(t) + \mathbf{F}_{\mathbf{u}}[\mathbf{u}(T-t), \mathbf{p}_F, T-t]\mathbf{d}(t) + \mathbf{V}\mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)] = \mathbf{0}, \quad (183)$$

$$\mathbf{d}(0) = -\mathbf{V}\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T)]. \quad (184)$$

For linear $\mathbf{F} = \mathbf{K}(t)\mathbf{u}$, the device propagates $\mathbf{d}$ in one additional finite-amplitude run with applied drive $-\mathbf{V}\mathbb{P}\theta_{\mathbf{u}}$, initial condition (184), and any explicit schedule reversed. The force-gradient cancellation (182) applies here as well.

For nonlinear trajectories, Onsager reciprocity does not remove the time-reversal obstruction of Sec. 6.2. At a linearly exponentially stable fixed point, the weighted nudge of Sec. 9.3 gives

$$\mathbf{F}[\mathbf{u}^{*\epsilon}, \mathbf{p}_F] - \mathbf{f} + \epsilon\mathbf{V}\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}^{*\epsilon}] = \mathbf{0}. \quad (185)$$

Linearizing at the free fixed point and using $\mathbf{V}$ -reciprocity shows that

$$\mathbf{d}^* := \lim_{\epsilon \rightarrow 0} \frac{\mathbf{u}^{*\epsilon} - \mathbf{u}^*}{\epsilon} = \mathbf{V}\mathbf{a}^*, \quad (186)$$

where $\mathbf{a}^*$ is the static adjoint (282). For $\mathbf{F} = \mathbf{V}E_{\mathbf{u}}$, the static force-gradient overlap likewise reduces to $\langle \mathbf{d}^*, E_{\mathbf{u}\mathbf{p}_F} \rangle$ when $\mathbf{V}$ is independent of $\mathbf{p}_F$.

For $\mathbf{F} = \mathbf{V}E_{\mathbf{u}}$ with constant $\mathbf{V} = \mathbf{V}^T \succ 0$ and constant drive $\mathbf{f}$, the free and nudged relaxations obey $\dot{\mathbf{u}} = -\mathbf{V}(E^\epsilon)_{\mathbf{u}}$ and descend the tilted energy in the $\mathbf{V}^{-1}$ metric:

$$\begin{aligned} E^\epsilon(\mathbf{u}) &= E(\mathbf{u}) - \mathbf{u}^T \mathbf{V}^{-1} \mathbf{f} + \epsilon \psi[\mathbb{P}\mathbf{u}], \\ \frac{dE^\epsilon}{dt} &= -(E^\epsilon)_{\mathbf{u}}^T \mathbf{V} (E^\epsilon)_{\mathbf{u}} \leq 0. \end{aligned} \quad (187)$$

A nondegenerate local minimum of E^ϵ is a linearly exponentially stable operating point. Both phases must still settle on the same stable branch.

D.2.1 Example: resistor–capacitor networks with heterogeneous capacitances

Let $\mathbf{u}(t)$ denote the node voltages of a passive resistor network with external current injections $\mathbf{j}(t)$. Kirchhoff's current law gives

$$\mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{G}\mathbf{u}(t) - \mathbf{j}(t) = \mathbf{0}, \quad \mathbf{C} = \mathbf{C}^T \succ 0, \quad \mathbf{G} = \mathbf{G}^T, \quad (188)$$

where $\mathbf{C}$ is the diagonal matrix of node capacitances and $\mathbf{G}$ is the conductance matrix. Define

$$\mathbf{V} := \mathbf{C}^{-1}, \quad \mathbf{K} := \mathbf{V}\mathbf{G}. \quad (189)$$

Then Eq. (188) becomes $\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} - \mathbf{V}\mathbf{j} = \mathbf{0}$. For heterogeneous capacitances, $\mathbf{K}$ is generally not symmetric, but it is $\mathbf{V}$ -reciprocal:

$$\mathbf{V}\mathbf{K}^T\mathbf{V}^{-1} = \mathbf{V}\mathbf{G}^T = \mathbf{V}\mathbf{G} = \mathbf{K}. \quad (190)$$

Write the distributed and terminal loss derivatives as

$$\mathbf{q}(t) := \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T-t)], \quad \mathbf{q}_T := \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T)]. \quad (191)$$

The transformed adjoint $\mathbf{d} = \mathbf{V}\mathbf{a}$ then obeys the circuit equations

$$\mathbf{C}\dot{\mathbf{d}}(t) + \mathbf{G}\mathbf{d}(t) + \mathbf{q}(t) = \mathbf{0}, \quad \mathbf{C}\mathbf{d}(0) = -\mathbf{q}_T. \quad (192)$$

Thus the physical adjoint run uses the current $\mathbf{j}^{(d)}(t) = -\mathbf{q}(t)$ and an initial capacitor-charge kick $-\mathbf{q}_T$. The circuit itself supplies the factors of $\mathbf{V}$. The Euclidean adjoint is recovered as $\mathbf{a} = \mathbf{V}^{-1}\mathbf{d} = \mathbf{C}\mathbf{d}$.

For the conductance gradients, hold the capacitances fixed and let g_e be the conductance of an edge $e = (m, n)$. With the incidence vector $\ell_e = \mathbf{e}_m - \mathbf{e}_n$ and $\mathbf{G}_{g_e} = \ell_e \ell_e^T$, Eq. (182) becomes

$$\begin{aligned} \frac{d\chi}{dg_e} &= \int_0^T \mathbf{a}(T-t)^T \mathbf{V} \mathbf{G}_{g_e} \mathbf{u}(t) dt \\ &= \int_0^T [d_m(T-t) - d_n(T-t)] [u_m(t) - u_n(t)] dt. \end{aligned} \quad (193)$$

Thus neither source preparation nor evaluation of the conductance gradients requires digital multiplication by $\mathbf{V}$. Gradients with respect to the capacitances must be treated separately.

The same metric applies to reciprocal nonlinear circuit elements whose current–voltage laws derive from a co-content function $\mathcal{E}(\mathbf{u})$ [46]:

$$\mathbf{C}\dot{\mathbf{u}} + \mathcal{E}_{\mathbf{u}}(\mathbf{u}) - \mathbf{j} = \mathbf{0}, \quad \mathbf{F}(\mathbf{u}) = \mathbf{V}\mathcal{E}_{\mathbf{u}}(\mathbf{u}). \quad (194)$$

Under the fixed-point assumptions above, the $\mathbf{V}$ -weighted nudge (185) is implemented by changing the applied current to $\mathbf{j}^\epsilon = \mathbf{j} - \epsilon \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}^{*,\epsilon}]$. This gives the Onsager-weighted Equilibrium Propagation construction. It does not provide a same-device adjoint for nonlinear first-order trajectories.

E Adjoint state equation and gradient for first-order systems

We derive the first-order adjoint equation (36) and the associated gradient formula quoted in Sec. 6. The derivation follows Appendix A step for step: the Lagrangian is again a functional of independent arguments, and each of its partial derivatives supplies one ingredient of the method. Stationarity in the multipliers returns the forward problem, stationarity in the state defines the adjoint problem, and the remaining explicit parameter dependence is the gradient. Two things are simpler here. The first-order problem carries a single initial condition, hence a single initial-condition multiplier, and its constraint integral needs only one integration by parts. We reuse the differentiability assumptions of Appendix A.

E.1 Constrained optimization problem

The forward dynamics is the first-order system (35),

$$\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, \quad t \in [0, T], \quad \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}), \quad (195)$$

with $\mathbf{u} \in C^1([0, T], V)$. The cost has trajectory and terminal parts. The terminal part depends on $\mathbf{u}(T)$ alone, since the full state of a first-order system is the position itself:

$$\chi[\mathbf{u}] = \int_0^T \theta[\mathbb{P}\mathbf{u}(t)] dt + \psi[\mathbb{P}\mathbf{u}(T)]. \quad (196)$$

The cost evaluated on the forward solution defines the reduced cost $\chi(\mathbf{p}) := \chi[\mathbf{u}(\cdot; \mathbf{p})]$, and the optimization problem is the following.

Problem E.1 (Constrained optimization of a first-order ODE). Find

$$\min_{\mathbf{p} \in \mathbb{R}^M} \chi[\mathbf{u}] \quad \text{subject to} \quad \begin{cases} \dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, & t \in [0, T], \\ \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}). \end{cases} \quad (197)$$

E.2 The Lagrangian

We enforce the state equation with a multiplier path $\mathbf{b} \in C^1([0, T], V)$ and the initial condition with a multiplier vector $\boldsymbol{\lambda} \in V$, and define

$$\mathcal{L}[\mathbf{u}, \mathbf{b}, \boldsymbol{\lambda}, \mathbf{p}] := \int_0^T \langle \mathbf{b}(t), \dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) \rangle dt + \langle \boldsymbol{\lambda}, \mathbf{u}(0) - \mathbf{u}_0(\mathbf{p}_{u_0}) \rangle + \chi[\mathbf{u}]. \quad (198)$$

As in Appendix A, the four arguments of $\mathcal{L}$ are *independent*: $\mathbf{u}$ is an arbitrary path, *not* assumed to solve the state equation, and $\mathbf{p}$ enters only through the explicit dependencies $\mathbf{F}[\cdot, \mathbf{p}_F, \cdot]$, $\mathbf{f}(\cdot, \mathbf{p}_f)$, $\mathbf{u}_0(\mathbf{p}_{u_0})$. We may therefore differentiate with respect to each argument separately, and the derivation consists of computing the four partial derivatives $\mathcal{L}_{\mathbf{b}}$, $\mathcal{L}_{\boldsymbol{\lambda}}$, $\mathcal{L}_{\mathbf{u}}$, $\mathcal{L}_{\mathbf{p}}$ in turn.

E.3 Stationarity with respect to the multipliers: the forward problem

The Lagrangian is affine in $\mathbf{b}$ and $\boldsymbol{\lambda}$, so both derivatives can be read off directly. Identifying them with elements of the respective spaces through the inner product,

$$\mathcal{L}_{\mathbf{b}}(t) = \dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f), \quad \mathcal{L}_{\boldsymbol{\lambda}} = \mathbf{u}(0) - \mathbf{u}_0(\mathbf{p}_{u_0}), \quad (199)$$

so $\mathcal{L}_{\mathbf{b}} = 0$ holds if and only if $\mathbf{u}$ satisfies the state equation (195) on $[0, T]$, and $\mathcal{L}_{\boldsymbol{\lambda}} = 0$ if and only if $\mathbf{u}$ satisfies the initial condition. Stationarity with respect to a multiplier again recovers the constraint that the multiplier enforces. Together the two conditions state that $\mathbf{u} = \mathbf{u}(\cdot; \mathbf{p})$ is the forward solution.

E.4 Stationarity turns the partial derivative into the gradient

Evaluated at the forward solution, the constraint terms in (198) vanish, so $\chi(\mathbf{p}) = \mathcal{L}[\mathbf{u}(\cdot; \mathbf{p}), \mathbf{b}, \boldsymbol{\lambda}, \mathbf{p}]$ for every $\mathbf{b}$ and $\boldsymbol{\lambda}$. Letting the multipliers also depend differentiably on $\mathbf{p}$ and differentiating gives one term per argument of $\mathcal{L}$:

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{u}} \frac{d\mathbf{u}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{b}} \frac{d\mathbf{b}}{d\mathbf{p}} + \mathcal{L}_{\boldsymbol{\lambda}} \frac{d\boldsymbol{\lambda}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{p}}. \quad (200)$$

The two multiplier terms vanish by Sec. E.3, whatever the sensitivities $d\mathbf{b}/d\mathbf{p}$ and $d\boldsymbol{\lambda}/d\mathbf{p}$ may be. The expensive sensitivity $d\mathbf{u}/d\mathbf{p}$ is multiplied only by $\mathcal{L}_{\mathbf{u}}$, while $\mathbf{b}$ and $\boldsymbol{\lambda}$ are still free. Choosing them such that $\mathcal{L}_{\mathbf{u}} = 0$ therefore leaves the identity

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{p}}, \quad (201)$$

in which only the cheap explicit parameter dependence of $\mathbf{F}$, $\mathbf{f}$ and $\mathbf{u}_0$ appears.

E.5 Stationarity with respect to the state: the adjoint problem

Let $\mathbf{h} \in C^1([0, T], V)$ be an arbitrary test path with no constraints at either endpoint. Varying $\mathbf{u} \rightarrow \mathbf{u} + \varepsilon \mathbf{h}$, and noting that the drive $\mathbf{f}$ has no $\mathbf{u}$ -dependence and does not contribute,

$$\mathcal{L}_{\mathbf{u}} \mathbf{h} = \int_0^T \langle \mathbf{b}, \dot{\mathbf{h}} + \mathbf{F}_{\mathbf{u}} \mathbf{h} \rangle dt + \langle \boldsymbol{\lambda}, \mathbf{h}(0) \rangle + \int_0^T \langle \mathbb{P} \theta_{\mathbf{u}}, \mathbf{h} \rangle dt + \langle \mathbb{P} \psi_{\mathbf{u}}, \mathbf{h}(T) \rangle. \quad (202)$$

The second-order case needed two integrations by parts; here, a single integration moves the derivative off $\mathbf{h}$:

$$\int_0^T \langle \mathbf{b}, \dot{\mathbf{h}} \rangle dt = - \int_0^T \langle \dot{\mathbf{b}}, \mathbf{h} \rangle dt + \langle \mathbf{b}(T), \mathbf{h}(T) \rangle - \langle \mathbf{b}(0), \mathbf{h}(0) \rangle, \quad (203)$$

and moving $\mathbf{F}_{\mathbf{u}}$ across the inner product gives $\langle \mathbf{b}, \mathbf{F}_{\mathbf{u}} \mathbf{h} \rangle = \langle \mathbf{F}_{\mathbf{u}}^T \mathbf{b}, \mathbf{h} \rangle$. Grouping by where $\mathbf{h}$ is evaluated,

$$\mathcal{L}_{\mathbf{u}} \mathbf{h} = \int_0^T \langle -\dot{\mathbf{b}} + \mathbf{F}_{\mathbf{u}}^T \mathbf{b} + \mathbb{P} \theta_{\mathbf{u}}, \mathbf{h} \rangle dt + \langle \mathbf{b}(T) + \mathbb{P} \psi_{\mathbf{u}}, \mathbf{h}(T) \rangle + \langle -\mathbf{b}(0) + \boldsymbol{\lambda}, \mathbf{h}(0) \rangle. \quad (204)$$

Since $\mathbf{h}$ is arbitrary and unconstrained at both endpoints, $\mathcal{L}_{\mathbf{u}} = 0$ requires each inner product to vanish separately. The bulk term and the term at $t = T$ involve only $\mathbf{b}$. Together they form a terminal-value problem that determines $\mathbf{b}$ uniquely and defines the adjoint state.

Definition E.2 (Adjoint state for the first-order problem). The *adjoint state* $\mathbf{b}(t) \in V$ associated with Problem E.1 is the solution of the adjoint ODE

$$-\dot{\mathbf{b}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(t), \mathbf{p}_F, t] \mathbf{b}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(t)] = \mathbf{0}, \quad t \in [0, T], \quad (205)$$

with the single required terminal condition

$$\mathbf{b}(T) = -\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T)]. \quad (206)$$

We solve the adjoint ODE (205) backward from T to 0 using this terminal condition. This produces $\mathbf{b}(0)$, which in turn determines the multiplier via the term at $t = 0$ in (204),

$$\boldsymbol{\lambda} = \mathbf{b}(0). \quad (207)$$

The adjoint ODE is linear in $\mathbf{b}$, its coefficients depend continuously on $\mathbf{p}$, and (207) is explicit. Hence $(\mathbf{b}, \boldsymbol{\lambda})$ is the unique solution of $\mathcal{L}_{\mathbf{u}} = 0$ and depends smoothly on $\mathbf{p}$, as required in Sec. E.4.

E.6 The gradient

Both remaining partial derivatives are now in place, and evaluating $\mathcal{L}_{\mathbf{p}}$ gives the gradient.

Theorem E.3 (Adjoint gradient for the first-order problem). *Let $\mathbf{u} = \mathbf{u}(\cdot; \mathbf{p})$ solve the forward problem of Problem E.1 and let $\mathbf{b}$ be the adjoint state of Definition E.2. Then the gradient of the reduced cost with respect to $\mathbf{p}$ is*

$$\frac{d\chi}{d\mathbf{p}} = \int_0^T \langle \mathbf{b}(t), \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}_f) \rangle dt - \langle \mathbf{b}(0), \frac{d\mathbf{u}_0}{d\mathbf{p}_{u_0}} \rangle. \quad (208)$$

Proof. With $(\mathbf{b}, \boldsymbol{\lambda})$ chosen as in Definition E.2 and Eq. (207), we have $\mathcal{L}_{\mathbf{u}} = 0$, and the key identity (201) gives $d\chi/d\mathbf{p} = \mathcal{L}_{\mathbf{p}}$. The explicit $\mathbf{p}$ -dependence of $\mathcal{L}$ enters through $\mathbf{F}[\mathbf{u}, \mathbf{p}_F, t]$ and $\mathbf{f}(t, \mathbf{p}_f)$ in the constraint integral and through $\mathbf{u}_0(\mathbf{p}_{u_0})$ in the initial-condition term. The cost χ has none. Differentiating each at fixed $\mathbf{u}$, $\mathbf{b}$, $\boldsymbol{\lambda}$ and substituting $\boldsymbol{\lambda} = \mathbf{b}(0)$ gives (208). $\square$

As in the second-order case, every quantity on the right-hand side is known: $\mathbf{u}(t)$ from the forward solve, $\mathbf{b}(t)$ from the backward solve, and the parameter sensitivities from the model. The computational cost is one forward solve plus one backward solve, independent of the number of parameters M .

Direct optimization of the initial condition. Suppose a subvector of $\mathbf{p}$ encodes the initial state itself, i.e. $\mathbf{u}_0(\mathbf{p}_{u_0}) = \mathbf{p}_{u_0}$ with $\mathbf{p}_{u_0} \in V$. Then $d\mathbf{u}_0/d\mathbf{p}_{u_0} = I_V$, and the $\mathbf{u}_0$ -block of Eq. (208) identifies the gradient directly:

$$\frac{d\chi}{d\mathbf{u}_0} = -\mathbf{b}(0), \quad (209)$$

that is, minus the initial-condition multiplier, which is the $\mathbf{M} = \mathbf{0}$, $\mathbf{D} = \mathbf{I}$ case of Eq. (142).

E.7 Integrating the adjoint forward in time

The adjoint problem of Definition E.2 is posed backward in time, while physical hardware runs forward. As in Sec. A.10, the change of variable $t \mapsto T - t$ removes the mismatch. Define the forward-time adjoint field

$$\mathbf{a}(t) := \mathbf{b}(T - t), \quad \dot{\mathbf{a}}(t) = -\dot{\mathbf{b}}(T - t). \quad (210)$$

Evaluating the adjoint ODE (205) at time $T - t$ and substituting (210) gives the forward-time adjoint equation

$$\dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}(T - t), \mathbf{p}_F, T - t] \mathbf{a}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T - t)] = \mathbf{0}, \quad t \in [0, T]. \quad (211)$$

Under the reversal the derivative term changes sign, $-\dot{\mathbf{b}} \rightarrow +\dot{\mathbf{a}}$. This is the $\mathbf{M} = \mathbf{0}$, $\mathbf{D} = \mathbf{I}$ instance of the flip $-\mathbf{D}^T \rightarrow +\mathbf{D}^T$ of Sec. A.10, and it gives Eq. (211) the same sign pattern as the state equation (195). Apart from this, the reversal only means that the linearization and the source are read along the reversed trajectory $\mathbf{u}(T - t)$, with the explicit time dependence of $\mathbf{F}$ replayed as $t \mapsto T - t$. The terminal condition (206) becomes an initial condition at the reversed time $t = 0$:

$$\mathbf{a}(0) = -\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T)]. \quad (212)$$

Likewise $\mathbf{b}(0) = \mathbf{a}(T)$, so substituting $\mathbf{b}(t) = \mathbf{a}(T - t)$ throughout the gradient (208) gives

$$\frac{d\chi}{d\mathbf{p}} = \int_0^T \langle \mathbf{a}(T - t), \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \mathbf{p}_F, t] - \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}_f) \rangle dt - \langle \mathbf{a}(T), \frac{d\mathbf{u}_0}{d\mathbf{p}_{u_0}} \rangle. \quad (213)$$

Eqs. (211) and (212) are the adjoint equation (36) with the source and initial condition (37) quoted in the main text. They are likewise the first-order analogues of the second-order results of Appendix A, and follow formally from them by the overdamped reduction $\mathbf{M} = \mathbf{0}$, $\mathbf{D} = \mathbf{I}$.

F Wirtinger derivation of the Schrödinger adjoint

We derive the Wirtinger adjoint state, gradient, and forward-time conjugated field used in Sec. 7 from an augmented Lagrangian in which $\mathbf{u}$ and $\bar{\mathbf{u}}$ are treated as independent variables [107]. The derivation parallels the real first-order case of Appendix E. The only new ingredient is Wirtinger calculus [86]. Wirtinger variables are natural here because real-valued objectives of complex fields are generically non-holomorphic: the variation of the cost depends on both $\delta\mathbf{u}$ and $\delta\bar{\mathbf{u}}$. This viewpoint is standard in complex optimization, where one treats a complex variable and its conjugate as formally independent variables to obtain stationarity conditions and descent directions [66–68]. For time-dependent complex dynamics, the same variational structure appears in quantum optimal control, where the Schrödinger equation is imposed by a complex Lagrange multiplier and stationarity of the augmented functional yields a terminal-value adjoint equation [64]. A rigorous PDE-constrained optimization formulation in complex Banach spaces, including semigroup Schrödinger equations and first- and second-order optimality conditions, was developed in [65]. Our derivation follows this complex Lagrangian/Wirtinger viewpoint, but is specialized to the adjoint field and forward-adjoint overlap gradients needed for the physical backpropagation constructions in Sec. 7.

F.1 Problem and conventions

The state obeys the Schrödinger-type system (41),

$$i\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, \quad \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}), \quad (214)$$

with $\mathbf{u} \in \mathbb{C}^N$, real parameters $\mathbf{p}$, and a real-valued cost

$$\chi[\mathbf{u}] = \int_0^T \theta[\mathbb{P}\mathbf{u}(t), \mathbb{P}\bar{\mathbf{u}}(t)] dt + \psi[\mathbb{P}\mathbf{u}(T), \mathbb{P}\bar{\mathbf{u}}(T)]. \quad (215)$$

We use the Hermitian inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{C}} = \bar{\mathbf{x}}^T \mathbf{y}$, treat $\mathbf{u}$ and $\bar{\mathbf{u}}$ as independent, and write the two Wirtinger Jacobians $\mathbf{F}_{\mathbf{u}}$ and $\mathbf{F}_{\bar{\mathbf{u}}}$. Since θ and ψ are real, $\theta_{\mathbf{u}} = \bar{\theta}_{\bar{\mathbf{u}}}$ and $\psi_{\mathbf{u}} = \bar{\psi}_{\bar{\mathbf{u}}}$, and, for an arbitrary variation $\mathbf{h}$ of $\mathbf{u}$, the first variation of the (real) cost is

$$\delta\chi = 2 \operatorname{Re} \left[\int_0^T \langle \mathbb{P}\theta_{\bar{\mathbf{u}}}, \mathbf{h} \rangle_{\mathbb{C}} dt + \langle \mathbb{P}\psi_{\bar{\mathbf{u}}}, \mathbf{h}(T) \rangle_{\mathbb{C}} \right]. \quad (216)$$

Problem F.1 (Constrained optimization of a Schrödinger-type system). Find

$$\min_{\mathbf{p}} \chi[\mathbf{u}] \quad \text{subject to} \quad \begin{cases} i\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f) = \mathbf{0}, & t \in [0, T], \\ \mathbf{u}(0) = \mathbf{u}_0(\mathbf{p}_{u_0}). \end{cases} \quad (217)$$

F.2 Augmented Lagrangian and the Wirtinger adjoint

Introduce a complex adjoint field $\mathbf{b} \in C^1([0, T], \mathbb{C}^N)$ and an initial-condition multiplier $\boldsymbol{\lambda} \in \mathbb{C}^N$, and pair the complex constraint against them through $2 \operatorname{Re}$ so that the Lagrangian is real,

$$\mathcal{L} = \chi + 2 \operatorname{Re} \int_0^T \langle \mathbf{b}(t), i\dot{\mathbf{u}} + \mathbf{F} - \mathbf{f} \rangle_{\mathbb{C}} dt + 2 \operatorname{Re} \langle \boldsymbol{\lambda}, \mathbf{u}(0) - \mathbf{u}_0 \rangle_{\mathbb{C}}. \quad (218)$$

The Lagrangian is affine in $\mathbf{b}$ and $\boldsymbol{\lambda}$, which enter only through the real pairing, so those two derivatives can be read off as the residuals they multiply,

$$\mathcal{L}_{\mathbf{b}}(t) = i\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t] - \mathbf{f}(t, \mathbf{p}_f), \quad \mathcal{L}_{\boldsymbol{\lambda}} = \mathbf{u}(0) - \mathbf{u}_0(\mathbf{p}_{u_0}). \quad (219)$$

A real pairing $2 \operatorname{Re} \langle \cdot, \mathbf{X} \rangle_{\mathbb{C}}$ that vanishes for every complex direction forces $\mathbf{X} = \mathbf{0}$, since a direction and its multiple by i recover the real and imaginary parts separately. Hence $\mathcal{L}_{\mathbf{b}} = 0$ holds if and only if $\mathbf{u}$ satisfies the state equation (214) on $[0, T]$, and $\mathcal{L}_{\boldsymbol{\lambda}} = 0$ if and only if it satisfies the initial condition. Stationarity in the multipliers returns the forward problem, as in the real case.

Evaluated on that forward solution, the two constraint pairings vanish and $\mathcal{L} = \chi(\mathbf{p})$. Letting the multipliers also depend differentiably on $\mathbf{p}$ and differentiating gives one term per argument of $\mathcal{L}$, with each product denoting the real pairing of a derivative with a sensitivity:

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{u}} \frac{d\mathbf{u}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{b}} \frac{d\mathbf{b}}{d\mathbf{p}} + \mathcal{L}_{\boldsymbol{\lambda}} \frac{d\boldsymbol{\lambda}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{p}}. \quad (220)$$

Here $\mathcal{L}_{\mathbf{u}}$ is the state derivative: its real pairing with a variation $\mathbf{h}$ gives the first variation computed below. Because that pairing is taken under $2 \operatorname{Re}$, it includes both the $\mathbf{u}$ and the $\bar{\mathbf{u}}$ contributions, and no separate condition in $\bar{\mathbf{u}}$ is needed. The two multiplier terms vanish by the previous paragraph, whatever the sensitivities $d\mathbf{b}/d\mathbf{p}$ and $d\boldsymbol{\lambda}/d\mathbf{p}$ may be, and the expensive sensitivity $d\mathbf{u}/d\mathbf{p}$ is multiplied only by $\mathcal{L}_{\mathbf{u}}$. Choosing $\mathbf{b}$ and $\boldsymbol{\lambda}$ so that $\mathcal{L}_{\mathbf{u}} = 0$ therefore leaves the key identity

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{p}}. \quad (221)$$

To impose stationarity of $\mathcal{L}$, we vary $\mathbf{u} \rightarrow \mathbf{u} + \varepsilon \mathbf{h}$ (with the accompanying conjugate variation $\bar{\mathbf{h}}$). The first variation $\delta\mathcal{L}$ is the coefficient of ε . The linearized residual is $i\dot{\mathbf{h}} + \mathbf{F}_{\mathbf{u}}\mathbf{h} + \mathbf{F}_{\bar{\mathbf{u}}}\bar{\mathbf{h}}$. Combining this with the cost variation (216) gives

$$\delta\mathcal{L} = 2 \operatorname{Re} \int_0^T \langle \mathbf{b}, i\dot{\mathbf{h}} + \mathbf{F}_{\mathbf{u}}\mathbf{h} + \mathbf{F}_{\bar{\mathbf{u}}}\bar{\mathbf{h}} \rangle_{\mathbb{C}} dt + 2 \operatorname{Re} \langle \boldsymbol{\lambda}, \mathbf{h}(0) \rangle_{\mathbb{C}} + \delta\chi. \quad (222)$$

To read off the stationarity conditions we move every term onto $\mathbf{h}$. The $\bar{\mathbf{h}}$ contribution is folded onto $\mathbf{h}$ under the real part,

$$2 \operatorname{Re} \langle \mathbf{b}, \mathbf{B} \bar{\mathbf{h}} \rangle_{\mathbb{C}} = 2 \operatorname{Re} \langle \mathbf{B}^T \bar{\mathbf{b}}, \mathbf{h} \rangle_{\mathbb{C}}, \quad (223)$$

applied with $\mathbf{B} = \mathbf{F}_{\bar{\mathbf{u}}}$, while the holomorphic term moves through the Hermitian adjoint, $2 \operatorname{Re} \langle \mathbf{b}, \mathbf{F}_{\mathbf{u}} \mathbf{h} \rangle_{\mathbb{C}} = 2 \operatorname{Re} \langle \bar{\mathbf{F}}_{\mathbf{u}}^T \bar{\mathbf{b}}, \mathbf{h} \rangle_{\mathbb{C}}$. The only time derivative is integrated by parts. Writing $\langle \mathbf{b}, i\dot{\mathbf{h}} \rangle_{\mathbb{C}} = \langle -i\mathbf{b}, \dot{\mathbf{h}} \rangle_{\mathbb{C}}$,

$$2 \operatorname{Re} \int_0^T \langle \mathbf{b}, i\dot{\mathbf{h}} \rangle_{\mathbb{C}} dt = 2 \operatorname{Re} \int_0^T \langle i\dot{\mathbf{b}}, \mathbf{h} \rangle_{\mathbb{C}} dt + 2 \operatorname{Re} \left[\langle -i\mathbf{b}(T), \mathbf{h}(T) \rangle_{\mathbb{C}} - \langle -i\mathbf{b}(0), \mathbf{h}(0) \rangle_{\mathbb{C}} \right]. \quad (224)$$

Every surviving factor of i originates from this step. Substituting (216)–(224) into (222) and grouping by where $\mathbf{h}$ is evaluated,

$$\begin{aligned} \delta \mathcal{L} = 2 \operatorname{Re} \int_0^T \langle i\dot{\mathbf{b}} + \bar{\mathbf{F}}_{\mathbf{u}}^T \bar{\mathbf{b}} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{b}} + \mathbb{P}\theta_{\bar{\mathbf{u}}}, \mathbf{h} \rangle_{\mathbb{C}} dt \\ + 2 \operatorname{Re} \langle -i\mathbf{b}(T) + \mathbb{P}\psi_{\bar{\mathbf{u}}}, \mathbf{h}(T) \rangle_{\mathbb{C}} + 2 \operatorname{Re} \langle i\mathbf{b}(0) + \boldsymbol{\lambda}, \mathbf{h}(0) \rangle_{\mathbb{C}}. \end{aligned} \quad (225)$$

Since $\mathbf{h}$ is arbitrary and unconstrained at both endpoints, $\mathcal{L}_{\mathbf{u}} = 0$ requires each bracket to vanish separately. The $t = 0$ term fixes the multiplier $\boldsymbol{\lambda} = -i\mathbf{b}(0)$, while the bulk and $t = T$ terms define the adjoint state.

Definition F.2 (Wirtinger adjoint state). The *Wirtinger adjoint state* $\mathbf{b}(t) \in \mathbb{C}^N$ associated with Problem F.1 is the solution of the backward-time ODE

$$i\dot{\mathbf{b}}(t) + \bar{\mathbf{F}}_{\mathbf{u}}^T \bar{\mathbf{b}}(t) + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{b}}(t) + \mathbb{P}\theta_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}(t), \mathbb{P}\bar{\mathbf{u}}(t)] = \mathbf{0}, \quad (226)$$

with the Wirtinger Jacobians evaluated at $(\mathbf{u}(t), \bar{\mathbf{u}}(t))$, and the terminal condition

$$\mathbf{b}(T) = -i \mathbb{P}\psi_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}(T), \mathbb{P}\bar{\mathbf{u}}(T)]. \quad (227)$$

Theorem F.3 (Wirtinger adjoint gradient). *Let $\mathbf{u}$ solve Problem F.1 and $\mathbf{b}$ be the adjoint state of Definition F.2. For real parameters $\mathbf{p}$, the gradient of the cost is*

$$\frac{d\chi}{d\mathbf{p}} = 2 \operatorname{Re} \int_0^T \langle \mathbf{b}(t), \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \bar{\mathbf{u}}(t), \mathbf{p}_F, t] - \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}_f) \rangle_{\mathbb{C}} dt + 2 \operatorname{Re} \langle i\mathbf{b}(0), \mathbf{u}_{0, \mathbf{p}_{u_0}} \rangle_{\mathbb{C}}. \quad (228)$$

Proof. With $\mathbf{b}$ and $\boldsymbol{\lambda}$ chosen as in Definition F.2, we have $\mathcal{L}_{\mathbf{u}} = 0$, and the key identity (221) gives $d\chi/d\mathbf{p} = \mathcal{L}_{\mathbf{p}}$. The explicit parameter dependence of (218) enters only through $\mathbf{F}(\mathbf{p}_F)$ and $\mathbf{f}(\mathbf{p}_f)$ in the constraint pairing and through $\mathbf{u}_0(\mathbf{p}_{u_0})$ in the initial-condition pairing, giving $\mathcal{L}_{\mathbf{p}} = 2 \operatorname{Re} \int_0^T \langle \mathbf{b}, \mathbf{F}_{\mathbf{p}_F} - \mathbf{f}_{\mathbf{p}_f} \rangle_{\mathbb{C}} dt - 2 \operatorname{Re} \langle \boldsymbol{\lambda}, \mathbf{u}_{0, \mathbf{p}_{u_0}} \rangle_{\mathbb{C}}$. Substituting $\boldsymbol{\lambda} = -i\mathbf{b}(0)$ yields (228). The external drive enters the residual as $-\mathbf{f}$, so its gradient carries the opposite sign to the internal force, and the initial-condition term carries the explicit factor i inherited from $\boldsymbol{\lambda} = -i\mathbf{b}(0)$. $\square$

F.3 Forward-time conjugated adjoint

The adjoint of Definition F.2 runs backward. For a physical realization, we use the conjugated time-reversed field $\mathbf{c}(t) := \bar{\mathbf{b}}(T - t)$. Evaluating (226) at $T - t$, using $\dot{\mathbf{b}}(T - t) = -\dot{\bar{\mathbf{c}}}(t)$, and complex conjugating the entire equation gives the forward-time equation (42),

$$\begin{aligned} i\dot{\mathbf{c}}(t) + \mathbf{F}_{\bar{\mathbf{u}}}^T[\mathbf{u}(T - t), \bar{\mathbf{u}}(T - t), \mathbf{p}_F, T - t] \mathbf{c}(t) \\ + \bar{\mathbf{F}}_{\mathbf{u}}^T[\mathbf{u}(T - t), \bar{\mathbf{u}}(T - t), \mathbf{p}_F, T - t]^T \bar{\mathbf{c}}(t) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(T - t), \mathbb{P}\bar{\mathbf{u}}(T - t)] = \mathbf{0}, \end{aligned} \quad (229)$$

where we used $\overline{\theta_{\mathbf{u}}} = \theta_{\mathbf{u}}$, with initial condition

$$\mathbf{c}(0) = \overline{\mathbf{b}(T)} = i \mathbb{P} \psi_{\mathbf{u}}[\mathbb{P} \mathbf{u}(T), \mathbb{P} \overline{\mathbf{u}}(T)]. \quad (230)$$

Conjugating term by term flips the signs in front of the Wirtinger Jacobians, which is why the conjugated field $\mathbf{c}$, rather than $\mathbf{b}$, can be propagated on the same hardware. The raw adjoint is recovered by $\mathbf{b}(t) = \overline{\mathbf{c}(T-t)}$. Substituting into (228) gives the gradient overlaps quoted in Sec. 7 in terms of the measured field,

$$\frac{d\chi}{d\mathbf{p}_F} = 2 \operatorname{Re} \int_0^T \langle \overline{\mathbf{c}(T-t)}, \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}(t), \overline{\mathbf{u}}(t), \mathbf{p}_F, t] \rangle_{\mathbb{C}} dt, \quad (231)$$

$$\frac{d\chi}{d\mathbf{p}_f} = -2 \operatorname{Re} \int_0^T \langle \overline{\mathbf{c}(T-t)}, \mathbf{f}_{\mathbf{p}_f}(t, \mathbf{p}_f) \rangle_{\mathbb{C}} dt, \quad (232)$$

$$\frac{d\chi}{d\mathbf{p}_{u_0}} = 2 \operatorname{Re} \langle i \overline{\mathbf{c}(T)}, \mathbf{u}_{0, \mathbf{p}_{u_0}} \rangle_{\mathbb{C}}. \quad (233)$$

F.4 Proof of the nonlinear intertwined-adjoint construction

We prove the nonlinear part of Theorem 9.1. Let τ denote time in the reversed experiment and set $r = T - \tau$. Use the loss data $\mathbf{q}(\tau)$ and $\mathbf{q}_T$ from Eq. (78), and abbreviate the forward-trajectory Jacobians by

$$\mathbf{A}(r) := \mathbf{F}_{\mathbf{u}}[\mathbf{u}(r), \overline{\mathbf{u}}(r), \mathbf{p}_F, r], \quad \mathbf{B}(r) := \mathbf{F}_{\overline{\mathbf{u}}}[\mathbf{u}(r), \overline{\mathbf{u}}(r), \mathbf{p}_F, r]. \quad (234)$$

The forward-time conjugated adjoint (42) is

$$\begin{aligned} i\dot{\mathbf{c}} + \mathbf{A}(r)^T \mathbf{c} + \overline{\mathbf{B}(r)}^T \overline{\mathbf{c}} + \mathbf{q} &= \mathbf{0}, \\ \mathbf{c}(0) &= i\mathbf{q}_T. \end{aligned} \quad (235)$$

Set $\mathbf{d} = \mathcal{Q}\mathbf{S}\mathbf{c}$. Multiplying Eq. (235) by $\mathcal{Q}\mathbf{S}$ and using Eqs. (82) and (83) gives

$$\begin{aligned} i\dot{\mathbf{d}} + \mathcal{Q}\overline{\mathbf{A}(r)}\mathcal{Q}^{-1}\mathbf{d} + \mathcal{Q}\overline{\mathbf{B}(r)}\mathcal{Q}^{-1}\overline{\mathbf{d}} + \mathcal{Q}\mathbf{S}\mathbf{q} &= \mathbf{0}, \\ \mathbf{d}(0) &= i\mathcal{Q}\mathbf{S}\mathbf{q}_T. \end{aligned} \quad (236)$$

No time-reversal identity has been used in this step. The matrix $\mathbf{S}$ acts only on the adjoint tangent equation.

Condition (81) now supplies the TRM. Define

$$\mathbf{w}(\tau) := \mathcal{Q}\overline{\mathbf{u}}(r). \quad (237)$$

Evaluating the forward equation at r , complex conjugating it, and multiplying by $\mathcal{Q}$ gives

$$\begin{aligned} i\dot{\mathbf{w}}(\tau) + \mathbf{F}[\mathbf{w}(\tau), \overline{\mathbf{w}}(\tau), \mathbf{p}_F, r] - \mathcal{Q}\overline{\mathbf{f}(r, \mathbf{p}_f)} &= \mathbf{0}, \\ \mathbf{w}(0) &= \mathcal{Q}\overline{\mathbf{u}(T)}. \end{aligned} \quad (238)$$

The two Wirtinger variables are independent. Differentiate Eq. (81) first with respect to $\overline{\mathbf{u}}$ and then with respect to $\mathbf{u}$. Evaluated along the forward trajectory, the two derivatives are

$$\begin{aligned} \mathbf{F}_{\mathbf{u}}[\mathbf{w}(\tau), \overline{\mathbf{w}}(\tau), \mathbf{p}_F, r] \mathcal{Q} &= \mathcal{Q}\overline{\mathbf{A}(r)}, \\ \mathbf{F}_{\overline{\mathbf{u}}}[\mathbf{w}(\tau), \overline{\mathbf{w}}(\tau), \mathbf{p}_F, r] \overline{\mathcal{Q}} &= \overline{\mathcal{Q}\mathbf{B}(r)}. \end{aligned} \quad (239)$$

The force covariance must therefore hold on an operating domain rather than only at isolated points of one trajectory.

Run the same system with

$$\begin{aligned} i\dot{\mathbf{w}}^\epsilon(\tau) + \mathbf{F}[\mathbf{w}^\epsilon(\tau), \overline{\mathbf{w}^\epsilon(\tau)}, \mathbf{p}_F, r] \\ - [\mathcal{Q}\overline{\mathbf{f}(r, \mathbf{p}_f)} - \epsilon\mathcal{Q}\mathbf{S}\mathbf{q}(\tau)] = \mathbf{0}, \\ \mathbf{w}^\epsilon(0) = \mathbf{w}(0) + \epsilon i\mathcal{Q}\mathbf{S}\mathbf{q}_T. \end{aligned} \quad (240)$$

The applied nudge is $-\epsilon\mathcal{Q}\mathbf{S}\mathbf{q}$ because the external drive enters the residual with a minus sign.

Let $\delta\mathbf{w} := \mathbf{w}^\epsilon - \mathbf{w}$. Subtract the unnudged equation and expand $\mathbf{F}$ to first order. Continuous Wirtinger differentiability implies differentiable dependence on ϵ on every interval where the perturbed trajectory remains in the operating domain. Using Eq. (239),

$$\begin{aligned} i\delta\dot{\mathbf{w}} + \mathcal{Q}\overline{\mathbf{A}(r)}\mathcal{Q}^{-1}\delta\mathbf{w} + \mathcal{Q}\overline{\mathbf{B}(r)}\mathcal{Q}^{-1}\delta\overline{\mathbf{w}} + \epsilon\mathcal{Q}\mathbf{S}\mathbf{q} + o(\epsilon) = \mathbf{0}, \\ \delta\mathbf{w}(0) = \epsilon i\mathcal{Q}\mathbf{S}\mathbf{q}_T. \end{aligned} \quad (241)$$

Divide by ϵ and take the limit. The resulting initial-value problem is Eq. (236), so uniqueness gives

$$\mathbf{d}(\tau) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(\tau) - \mathbf{w}(\tau)}{\epsilon} = \mathcal{Q}\mathbf{S}\mathbf{c}(\tau), \quad \mathbf{c}(\tau) = \mathbf{S}^{-1}\mathcal{Q}^{-1}\mathbf{d}(\tau). \quad (242)$$

Thus both transformations are undone, in reverse order, when the adjoint is recovered from the measured response. $\square$

G Second-order systems: gyroscopic $\mathcal{PT}$ mirrors

For an undamped gyroscopic system, the background mirror and tangent matching can also be treated separately. Let $r := T - \tau$ and consider

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{G}\dot{\mathbf{u}} + \mathbf{F}[\mathbf{u}, \mathbf{p}_F, t] = \mathbf{f}(t), \quad \mathbf{M}^T = \mathbf{M}, \quad \mathbf{G}^T = -\mathbf{G}. \quad (243)$$

Gyroscopic terms describe Coriolis forces, Lorentz forces on charged lattices, and the dynamics of gyroscopic metamaterials [97]. Let $\mathbf{\Pi} = \mathbf{\Pi}^T = \mathbf{\Pi}^{-1}$ be a real spatial mirror. Suppose

$$\mathbf{\Pi}\mathbf{M}\mathbf{\Pi} = \mathbf{M}, \quad \mathbf{\Pi}\mathbf{G}\mathbf{\Pi} = -\mathbf{G}, \quad \mathbf{F}[\mathbf{\Pi}\mathbf{u}(t), \mathbf{p}_F, t] = \mathbf{\Pi}\mathbf{F}[\mathbf{u}(t), \mathbf{p}_F, t] \quad (244)$$

along the forward trajectory. Then

$$\mathbf{w}(\tau) = \mathbf{\Pi}\mathbf{u}(r), \quad \mathbf{w}(0) = \mathbf{\Pi}\mathbf{u}(T), \quad \dot{\mathbf{w}}(0) = -\mathbf{\Pi}\dot{\mathbf{u}}(T) \quad (245)$$

is generated on the same hardware by reversing the explicit schedule and applying the drive $\mathbf{\Pi}\mathbf{f}(r)$.

Let

$$\mathbf{J}(t) := \mathbf{F}_{\mathbf{u}}[\mathbf{u}(t), \mathbf{p}_F, t], \quad \tilde{\mathbf{J}}(t) := \mathbf{F}_{\mathbf{u}}[\mathbf{\Pi}\mathbf{u}(t), \mathbf{p}_F, t].$$

With no additional tangent intertwiner, the required matching condition is

$$\mathbf{\Pi}\tilde{\mathbf{J}}(t)\mathbf{\Pi} = \mathbf{J}(t)^T. \quad (246)$$

If the force covariance in Eq. (244) holds throughout an invariant operating domain, differentiating it gives $\tilde{\mathbf{J}} = \mathbf{\Pi}\mathbf{J}\mathbf{\Pi}$, and Eq. (246) reduces to $\mathbf{J}^T = \mathbf{J}$.

Define the adjoint source locally by

$$\mathbf{f}_{\text{adj}}(\tau) = \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(r)].$$

The nudged drive is

$$\mathbf{f}^\epsilon(\tau) = \mathbf{\Pi}\mathbf{f}(r) - \epsilon\mathbf{\Pi}\mathbf{f}_{\text{adj}}(\tau).$$

The endpoint perturbations are

$$\mathbf{w}^\epsilon(0) = \mathbf{w}(0) + \epsilon \mathbf{\Pi} \mathbf{a}(0), \quad \dot{\mathbf{w}}^\epsilon(0) = \dot{\mathbf{w}}(0) + \epsilon \mathbf{\Pi} \dot{\mathbf{a}}(0),$$

where the gyroscopic adjoint data satisfy

$$\mathbf{M} \mathbf{a}(0) = -\mathbb{P} \psi_{\dot{\mathbf{u}}}, \quad \mathbf{M} \dot{\mathbf{a}}(0) = \mathbf{G} \mathbf{a}(0) - \mathbb{P} \psi_{\mathbf{u}}. \quad (247)$$

The loss derivatives are evaluated at the terminal forward state. The differential response is

$$\mathbf{d}(\tau) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{w}^\epsilon(\tau) - \mathbf{w}(\tau)}{\epsilon} = \mathbf{\Pi} \mathbf{a}(\tau), \quad \mathbf{a} = \mathbf{\Pi} \mathbf{d}. \quad (248)$$

This is a separate extension of the second-order adjoint derivation, not an application of the nonlinear branch of Theorem 5.1.³

An additional tangent intertwiner can be included by setting $\mathbf{d} = \mathbf{\Pi} \mathbf{S} \mathbf{a}$, where $\mathbf{S}$ is known, real, invertible, and constant during the experiment. Under force covariance throughout an invariant operating domain, the following conditions must hold for all t along the forward trajectory and throughout the visited trainable family:

$$\mathbf{S} \mathbf{M}^T \mathbf{S}^{-1} = \mathbf{M}, \quad \mathbf{S} \mathbf{G}^T \mathbf{S}^{-1} = -\mathbf{G}, \quad \mathbf{S} \mathbf{J}^T \mathbf{S}^{-1} = \mathbf{J}.$$

The corresponding drive and endpoint perturbations are

$$\mathbf{f}^\epsilon(\tau) = \mathbf{\Pi} \mathbf{f}(r) - \epsilon \mathbf{\Pi} \mathbf{S} \mathbf{f}_{\text{adj}}(\tau),$$

$$\mathbf{w}^\epsilon(0) = \mathbf{w}(0) + \epsilon \mathbf{\Pi} \mathbf{S} \mathbf{a}(0), \quad \dot{\mathbf{w}}^\epsilon(0) = \dot{\mathbf{w}}(0) + \epsilon \mathbf{\Pi} \mathbf{S} \dot{\mathbf{a}}(0).$$

The adjoint is recovered as $\mathbf{a} = \mathbf{S}^{-1} \mathbf{\Pi} \mathbf{d}$.

H Free-space optics

Free-space diffractive optics. In the scalar, monochromatic, paraxial, and thin-element approximation, a free-space diffractive optical system is a linear Schrödinger-type system of the form considered in Sec. 7, with the propagation coordinate z playing the role of time [109]. Starting from the scalar Helmholtz equation and writing the field as a slowly varying envelope,

$$(\nabla_\perp^2 + \partial_z^2 + k_0^2 n^2(\mathbf{r}_\perp, z)) E(\mathbf{r}_\perp, z) = 0, \quad E(\mathbf{r}_\perp, z) = u(\mathbf{r}_\perp, z) e^{ikz}, \quad k = k_0 n_0, \quad (249)$$

the paraxial approximation neglects $\partial_z^2 u$ relative to $2ik\partial_z u$, giving

$$i\partial_z u + \left[\frac{1}{2k} \nabla_\perp^2 + k_0 \Delta n(\mathbf{r}_\perp, z) \right] u = 0. \quad (250)$$

An ideal SLM plane at $z = z_j$ is modeled as a thin local transmission

$$u(z_j^+) = m_j(\mathbf{r}_\perp) u(z_j^-), \quad m_j(\mathbf{r}_\perp) = \rho_j(\mathbf{r}_\perp) e^{i\phi_j(\mathbf{r}_\perp)} = e^{i\alpha_j(\mathbf{r}_\perp)}, \quad (251)$$

or equivalently as a delta-function potential, the form underlying split-step beam propagation [110],

$$i\partial_z u + \left[\frac{1}{2k} \nabla_\perp^2 + \sum_j \alpha_j(\mathbf{r}_\perp) \delta(z - z_j) \right] u = 0. \quad (252)$$

³For the $\mathbf{\Pi}$ -only construction, a parity-odd velocity coupling $\mathbf{N}$ must satisfy both $\mathbf{\Pi} \mathbf{N} \mathbf{\Pi} = -\mathbf{N}$ to reverse the background and $\mathbf{\Pi} \mathbf{N}^T \mathbf{\Pi} = \mathbf{N}$ to match the adjoint. Together these imply $\mathbf{N}^T = -\mathbf{N}$. This excludes dissipative balanced gain and loss from the $\mathbf{\Pi}$ -only construction [108]; an additional intertwiner may admit special cases.

After transverse discretization this has the form treated in Sec. 7,

$$i\dot{\mathbf{u}}(z) + \mathbf{F}(\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F) = 0, \quad \mathbf{F}(\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F) = \mathbf{K}(z; \mathbf{p}_K)\mathbf{u}, \quad (253)$$

with

$$\mathbf{K}(z; \mathbf{p}_K) = \frac{1}{2k}\mathbf{L}_\perp + \sum_j \mathbf{A}_j(\mathbf{p}_j)\delta(z - z_j), \quad \mathbf{D}_j = e^{i\mathbf{A}_j} = \text{diag}(m_{j1}, \dots, m_{jN}). \quad (254)$$

The internal force (253) is linear and *holomorphic*: it does not depend on $\bar{\mathbf{u}}$, so $\mathbf{F}_{\bar{\mathbf{u}}} = \mathbf{0}$ and the second Wirtinger Jacobian drops out of every formula in Appendix F. Equivalently, the input-output map is the usual free-space/SLM cascade [11]

$$\mathbf{U} = \mathbf{P}_m \mathbf{D}_m \mathbf{P}_{m-1} \mathbf{D}_{m-1} \cdots \mathbf{P}_1 \mathbf{D}_1 \mathbf{P}_0, \quad \mathbf{P}_\ell = \exp\left(i \frac{\Delta z_\ell}{2k} \mathbf{L}_\perp\right). \quad (255)$$

For a reciprocal scalar discretization,

$$\mathbf{L}_\perp^T = \mathbf{L}_\perp, \quad \mathbf{A}_j^T = \mathbf{A}_j, \quad \mathbf{D}_j^T = \mathbf{D}_j. \quad (256)$$

Reciprocity therefore requires *complex symmetry*, not Hermiticity or unitarity. In particular, lossy or linear-gain pixels with $|m_{ji}| \neq 1$ are admissible even though generally

$$\bar{\mathbf{D}}_j^T \mathbf{D}_j \neq \mathbf{I}. \quad (257)$$

Accordingly, the physical adjoint experiment is the $\mathbf{c}$ -adjoint run of Appendix F, not a direct integration of the backward Wirtinger adjoint $\mathbf{b}$. In the linear reciprocal case the $\mathbf{c}$ -field obeys

$$i\dot{\mathbf{c}}(z) + \mathbf{K}^T(L - z; \mathbf{p}_K)\mathbf{c}(z) + \mathbb{P}\theta_{\mathbf{u}}[\mathbb{P}\mathbf{u}(L - z), \mathbb{P}\bar{\mathbf{u}}(L - z)] = 0, \quad (258)$$

with initial condition

$$\mathbf{c}(0) = i \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}(L), \mathbb{P}\bar{\mathbf{u}}(L)]. \quad (259)$$

Since $\mathbf{K}^T = \mathbf{K}$, this is implemented by the same reciprocal optical system traversed in the reverse direction, with the adjoint source and initial condition applied at the output side. The operator $\mathbf{K}^T(L - z)$ places the modulator planes at $L - z_j$, in the element order realized by reverse traversal. The backward adjoint is then recovered as

$$\mathbf{b}(z) = \overline{\mathbf{c}(L - z)}. \quad (260)$$

For parameters that enter only through the optical operator, the gradient of Theorem F.3 becomes

$$\frac{d\chi}{dp} = 2 \text{Re} \int_0^L \left\langle \overline{\mathbf{c}(L - z)}, \mathbf{K}_p(z; \mathbf{p}_K)\mathbf{u}(z) \right\rangle_{\mathbb{C}} dz. \quad (261)$$

Pixel gradients. Each SLM pixel carries two trainable parameters, a phase and a log-amplitude,

$$m_{ji} = e^{\gamma_{ji} + i\phi_{ji}} = \rho_{ji} e^{i\phi_{ji}}, \quad \rho_{ji} = e^{\gamma_{ji}}, \quad \alpha_{ji} = \phi_{ji} - i\gamma_{ji}, \quad (262)$$

so that $\gamma_{ji} < 0$ describes an attenuating pixel, $\gamma_{ji} > 0$ a linear-gain pixel, and a phase-only device is the special case $\gamma_{ji} \equiv 0$. Both gradients are obtained from the same pair of runs. We use the cascade to derive the gradient because it exhibits the pixel dependence directly [21]. Letting $\mathbf{E}_i$ be diagonal with a single nonzero entry at pixel i ,

$$\frac{\partial \mathbf{U}}{\partial \phi_{ji}} = \mathbf{P}_m \cdots \mathbf{P}_j (i m_{ji} \mathbf{E}_i) \mathbf{P}_{j-1} \cdots \mathbf{P}_0, \quad (263)$$

and since $\mathbf{E}_i \mathbf{D}_j = \mathbf{D}_j \mathbf{E}_i$ the phase and amplitude gradients are the two parts of one complex overlap at plane j ,

$$\frac{\partial \chi}{\partial \phi_{ji}} = 2 \operatorname{Re} [\overline{b_i} u_i]_{z_j^+}, \quad (264)$$

$$\frac{\partial \chi}{\partial \gamma_{ji}} = 2 \operatorname{Im} [\overline{b_i} u_i]_{z_j^+}. \quad (265)$$

Both fields must be evaluated on the *same* side of plane j , here the output side z_j^+ . The assignment of the real and imaginary parts follows from the factor $-i$ in $\mathbf{b}(L) = -i \mathbb{P} \psi_{\bar{\mathbf{u}}}$ of Definition F.2, which rotates the overlap by a quarter turn.

I Details for stationary problems

This appendix gives the derivations behind Sec. 8. The setting is the stationary problem

$$\mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F] - \mathbf{f}(\mathbf{p}_f) = \mathbf{0}, \quad (266)$$

with an invertible tangent operator at the solution. This includes static linear elasticity, linear resistor networks, and frequency-domain scattering at a fixed drive frequency. We keep the complex-valued notation throughout. Real problems are the special case treated in the last subsection. The loss is a real-valued function of the measured field,

$$\chi = \psi[\mathbb{P}\mathbf{u}, \mathbb{P}\bar{\mathbf{u}}]. \quad (267)$$

Problem I.1 (Constrained optimization of a stationary system). Find

$$\min_{\mathbf{p}} \psi[\mathbb{P}\mathbf{u}, \mathbb{P}\bar{\mathbf{u}}] \quad \text{subject to} \quad \mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F] = \mathbf{f}(\mathbf{p}_f), \quad (268)$$

over the parameters $\mathbf{p} = [\mathbf{p}_F, \mathbf{p}_f] \in \mathbb{R}^M$.

The derivation follows the preceding appendices: the field equation is imposed as a constraint with an adjoint multiplier. Stationarity with respect to the multiplier returns the field equation, while stationarity with respect to the field yields the adjoint equation. The remaining explicit parameter dependence gives the gradient as a forward–adjoint overlap. For frequency-domain problems this is the standard adjoint-variable sensitivity construction used for Maxwell and Helmholtz systems [63].

I.1 The stationary adjoint

There is no time, hence no integration by parts and no boundary term, and the single algebraic constraint carries a single multiplier. Pairing the residual through $2 \operatorname{Re}$ gives

$$\mathcal{L}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{a}, \bar{\mathbf{a}}, \mathbf{p}] = \chi + 2 \operatorname{Re} \langle \mathbf{a}, \mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F] - \mathbf{f}(\mathbf{p}_f) \rangle_{\mathbb{C}}, \quad (269)$$

a real-valued function of independent arguments in which $\mathbf{u}$ is an arbitrary field and is *not* assumed to solve (266). The partial derivatives of $\mathcal{L}$ below include the loss contribution.

Stationarity in the multiplier returns the constraint,

$$\mathcal{L}_{\mathbf{a}} = \mathbf{F}[\mathbf{u}, \bar{\mathbf{u}}, \mathbf{p}_F] - \mathbf{f}(\mathbf{p}_f), \quad (270)$$

so $\mathcal{L}_{\mathbf{a}} = 0$ holds if and only if $\mathbf{u}$ solves (266), since a real pairing that vanishes in every complex direction forces the residual to vanish, as in Appendix F. At the solution, the constraint term vanishes and $\mathcal{L} = \chi(\mathbf{p})$, so differentiating with one term per argument of $\mathcal{L}$ gives

$$\frac{d\chi}{d\mathbf{p}} = \mathcal{L}_{\mathbf{u}} \frac{d\mathbf{u}}{d\mathbf{p}} + \mathcal{L}_{\bar{\mathbf{u}}} \frac{d\bar{\mathbf{u}}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{a}} \frac{d\mathbf{a}}{d\mathbf{p}} + \mathcal{L}_{\bar{\mathbf{a}}} \frac{d\bar{\mathbf{a}}}{d\mathbf{p}} + \mathcal{L}_{\mathbf{p}}. \quad (271)$$

The two multiplier terms vanish by (270), whatever the sensitivities of $\mathbf{a}$ may be. Computing $d\mathbf{u}/d\mathbf{p}$ would otherwise require one stationary solve per parameter. It appears only multiplied by $\mathcal{L}_{\mathbf{u}}$ and its conjugate. Choosing $\mathbf{a}$ so that $\mathcal{L}_{\bar{\mathbf{u}}} = 0$, which by conjugacy also gives $\mathcal{L}_{\mathbf{u}} = 0$, therefore leaves the key identity $d\chi/d\mathbf{p} = \mathcal{L}_{\mathbf{p}}$, and that condition is what defines the adjoint. Writing the real pairing out,

$$2 \operatorname{Re} \langle \mathbf{a}, \mathbf{F} - \mathbf{f} \rangle_{\mathbb{C}} = \bar{\mathbf{a}}^T (\mathbf{F} - \mathbf{f}) + \mathbf{a}^T (\bar{\mathbf{F}} - \bar{\mathbf{f}}), \quad (272)$$

and collecting the terms that carry $\bar{\mathbf{u}}$, through the Wirtinger Jacobian $\mathbf{F}_{\bar{\mathbf{u}}}$ in the first term and $\bar{\mathbf{F}}_{\bar{\mathbf{u}}} = \bar{\mathbf{F}}_{\mathbf{u}}$ in the second,

$$\mathcal{L}_{\bar{\mathbf{u}}} = \mathbb{P}\psi_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}, \mathbb{P}\bar{\mathbf{u}}] + \bar{\mathbf{F}}_{\mathbf{u}}^T \mathbf{a} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{a}}. \quad (273)$$

Setting it to zero gives the adjoint.

Definition I.2 (Stationary adjoint state). The adjoint state $\mathbf{a} \in \mathbb{C}^N$ associated with Problem I.1 is the solution of $\mathcal{L}_{\bar{\mathbf{u}}} = 0$, that is

$$\bar{\mathbf{F}}_{\mathbf{u}}^T \mathbf{a} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{a}} = -\mathbb{P}\psi_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}, \mathbb{P}\bar{\mathbf{u}}], \quad (274)$$

with both Wirtinger Jacobians evaluated at the solution of (266).

Theorem I.3 (Stationary adjoint gradient). Let $\mathbf{u}$ solve Problem I.1 and $\mathbf{a}$ be the adjoint state of Definition I.2. Then, for each parameter p_m ,

$$\frac{d\chi}{dp_m} = 2 \operatorname{Re} \langle \mathbf{a}, \mathbf{F}_{p_m} - \mathbf{f}_{p_m} \rangle_{\mathbb{C}}. \quad (275)$$

Proof. With $\mathbf{a}$ as in Definition I.2 we have $\mathcal{L}_{\bar{\mathbf{u}}} = 0$ and hence $\mathcal{L}_{\mathbf{u}} = 0$, so (271) leaves $d\chi/d\mathbf{p} = \mathcal{L}_{\mathbf{p}}$, and the explicit parameter dependence of the Lagrangian (269) sits in $\mathbf{F}$ and $\mathbf{f}$ alone. $\square$

I.2 Linear systems

For $\mathbf{F} = \mathbf{K}(\mathbf{p}_{\mathbf{K}})\mathbf{u}$ with invertible $\mathbf{K}$, the Wirtinger Jacobians are $\mathbf{F}_{\mathbf{u}} = \mathbf{K}$ and $\mathbf{F}_{\bar{\mathbf{u}}} = \mathbf{0}$, and the adjoint equation (274) becomes the Hermitian-adjoint system

$$\mathbf{K}(\mathbf{p}_{\mathbf{K}})^{\dagger} \mathbf{a} = -\mathbb{P}\psi_{\bar{\mathbf{u}}}[\mathbb{P}\mathbf{u}, \mathbb{P}\bar{\mathbf{u}}], \quad (276)$$

and the gradients of Theorem I.3 are

$$\frac{d\chi}{d\mathbf{p}_{\mathbf{K}}} = 2 \operatorname{Re} \langle \mathbf{a}, \mathbf{K}_{\mathbf{p}_{\mathbf{K}}} \mathbf{u} \rangle_{\mathbb{C}}, \quad (277)$$

$$\frac{d\chi}{d\mathbf{p}_{\mathbf{f}}} = -2 \operatorname{Re} \langle \mathbf{a}, \mathbf{f}_{\mathbf{p}_{\mathbf{f}}} \rangle_{\mathbb{C}}. \quad (278)$$

The conjugated equation (274) becomes

$$\mathbf{K}(\mathbf{p}_{\mathbf{K}})^T \mathbf{c} = -\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}, \mathbb{P}\bar{\mathbf{u}}]. \quad (279)$$

For a reciprocal system, $\mathbf{K}^T = \mathbf{K}$, so $\mathbf{c}$ is obtained by solving the *same* forward problem with the source replaced by $-\mathbb{P}\psi_{\mathbf{u}}$, applied at the measurement degrees of freedom. In terms of that physically generated field the gradients read

$$\frac{d\chi}{d\mathbf{p}_K} = 2 \operatorname{Re}[\mathbf{c}^T \mathbf{K}_{\mathbf{p}_K} \mathbf{u}], \quad (280)$$

$$\frac{d\chi}{d\mathbf{p}_f} = -2 \operatorname{Re}[\mathbf{c}^T \mathbf{f}_{\mathbf{p}_f}]. \quad (281)$$

Eq. (279) is Eq. (68) of the main text, and (280)–(281) are the gradients (62)–(63) specialized to $\mathbf{F} = \mathbf{K}\mathbf{u}$.

I.3 Real fields: the static adjoint at a fixed point

When $\mathbf{F}$ is real and nonlinear, the trajectory-based construction is obstructed (Theorem 6.1 and Sec. 6.2). What remains accessible is the fixed point of the relaxation

$$\dot{\mathbf{u}}(t) + \mathbf{F}[\mathbf{u}(t), \mathbf{p}_F] - \mathbf{f}(\mathbf{p}_f) = \mathbf{0},$$

with a time-independent drive and a terminal-cost-only objective ($\theta \equiv 0$). We assume that the relaxation converges to a linearly exponentially stable fixed point $\mathbf{u}^*(\mathbf{p})$. The loss $\psi[\mathbb{P}\mathbf{u}^*]$ is measured after settling. The relevant parameters are those entering the fixed-point equation. Parameters affecting only the initial condition leave $\mathbf{u}^*$ unchanged as long as the trajectory stays in the same basin of attraction. For real fields the conjugations in the derivation above drop and the factor $2 \operatorname{Re}$ becomes the Euclidean inner product, so the adjoint equation (274) reduces to the following.

Definition I.4 (Static adjoint state). The *static adjoint* $\mathbf{a}^*(\mathbf{p})$ is the solution of

$$\mathbf{F}_{\mathbf{u}}^T[\mathbf{u}^*(\mathbf{p}), \mathbf{p}_F] \mathbf{a}^* + \mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}^*(\mathbf{p})] = \mathbf{0}. \quad (282)$$

No separate invertibility hypothesis is needed. Linearizing the forward dynamics about the fixed point gives $\delta\dot{\mathbf{u}} = -\mathbf{F}_{\mathbf{u}}\delta\mathbf{u}$, so linear exponential stability of $\mathbf{u}^*$ places the spectrum of $\mathbf{F}_{\mathbf{u}}[\mathbf{u}^*, \mathbf{p}_F]$ in the open right half-plane, $\operatorname{Re} \lambda(\mathbf{F}_{\mathbf{u}}) > 0$. In particular $\mathbf{F}_{\mathbf{u}}$ is invertible, $\mathbf{a}^*$ is the unique solution of (282), and $\mathbf{u}^*(\mathbf{p})$, $\mathbf{a}^*(\mathbf{p})$ depend smoothly on $\mathbf{p}$ by the implicit function theorem.

Theorem I.5 (Static adjoint gradient). Let $\mathbf{u}^*$ be the linearly exponentially stable fixed point above, so that $\mathbf{F}_{\mathbf{u}}[\mathbf{u}^*, \mathbf{p}_F]$ is invertible, and $\mathbf{a}^*$ the static adjoint of Definition I.4. Then

$$\frac{d\chi}{d\mathbf{p}} = \langle \mathbf{a}^*, \mathbf{F}_{\mathbf{p}_F}[\mathbf{u}^*(\mathbf{p}), \mathbf{p}_F] - \mathbf{f}_{\mathbf{p}_f}(\mathbf{p}_f) \rangle. \quad (283)$$

Proof. The real instance of Theorem I.3. □

Relation to the time-dependent adjoint. The static adjoint is the time integral of the forward-time adjoint of Appendix E. Take the forward system to be settled, $\mathbf{u}(t) \equiv \mathbf{u}^*$ on the whole window, which is the free phase of the protocol below. The linearization is then constant along the trajectory, and with $\theta \equiv 0$ the running source vanishes, so (211) is the homogeneous relaxation

$$\dot{\mathbf{a}}(t) + \mathbf{F}_{\mathbf{u}}^T[\mathbf{u}^*, \mathbf{p}_F] \mathbf{a}(t) = \mathbf{0}, \quad \mathbf{a}(0) = -\mathbb{P}\psi_{\mathbf{u}}[\mathbb{P}\mathbf{u}^*], \quad \mathbf{a}(t) = e^{-\mathbf{F}_{\mathbf{u}}^T t} \mathbf{a}(0). \quad (284)$$

The spectrum of $\mathbf{F}_{\mathbf{u}}$ lies in the open right half-plane, so $\mathbf{a}$ decays to zero and its late-time value carries no information. Its time integral does:

$$\int_0^\infty \mathbf{a}(t) dt = \mathbf{F}_{\mathbf{u}}^{-T} \mathbf{a}(0) = -\mathbf{F}_{\mathbf{u}}^{-T} \mathbb{P}\psi_{\mathbf{u}} = \mathbf{a}^*, \quad (285)$$

which is (282). Over a window long compared with the decay time, truncating the integral to $[0, T]$ makes no difference. The two gradient formulas agree for the same reason: on the settled trajectory the second factor in the integrand of (213) is the constant $\mathbf{F}_{\mathbf{p}_F}[\mathbf{u}^*, \mathbf{p}_F] - \mathbf{f}_{\mathbf{p}_f}(\mathbf{p}_f)$, so the time integral falls entirely on the adjoint and returns $\mathbf{a}^*$, recovering (283). The initial-condition term drops out because $\mathbf{a}(T)$ has decayed, matching the remark above that parameters entering only $\mathbf{u}_0$ leave $\mathbf{u}^*$ unchanged.

Physical realization: Equilibrium Propagation. The static adjoint (282) requires the transpose $\mathbf{F}_{\mathbf{u}}^T$, which is generally not realizable on the same device. Reciprocity removes it: for $\mathbf{F}_{\mathbf{u}}^T = \mathbf{F}_{\mathbf{u}}$, i.e. $\mathbf{F} = E_{\mathbf{u}}$, the static adjoint is the infinitesimal difference between the free fixed point and one nudged by the adjoint source, $\mathbf{a}^* = \lim_{\epsilon \rightarrow 0} (\mathbf{u}^{*\epsilon} - \mathbf{u}^*)/\epsilon$, the real-valued instance of Eq. (71). Both relaxations descend the energy E tilted by the drive, with the second additionally tilted by $\epsilon\psi$. This gives the two-phase energy contrast of Scellier and Bengio [29]. Substituting the difference of equilibria into Theorem I.5 gives the learning rule.

I.4 Layered photonic networks

Integrated photonic neural networks can be described as a sequence of frequency-domain optical solves and elementwise activations. For $\ell = 1, \dots, L$, write

$$\mathbf{K}_\ell(\mathbf{p})\mathbf{u}_\ell = \mathbf{B}_\ell \mathbf{X}_{\ell-1}, \quad (286a)$$

$$\mathbf{Z}_\ell = \mathbf{P}_\ell \mathbf{u}_\ell, \quad (286b)$$

$$\mathbf{X}_\ell = f_\ell(\mathbf{Z}_\ell, \overline{\mathbf{Z}}_\ell). \quad (286c)$$

Here $\mathbf{K}_\ell$ is the reciprocal frequency-domain operator of layer ℓ , $\mathbf{B}_\ell$ injects the incoming modal amplitudes into the device, and $\mathbf{P}_\ell$ projects the internal field onto output modes. Eliminating the field gives the usual layer transfer matrix

$$\mathbf{Z}_\ell = \widehat{\mathbf{W}}_\ell \mathbf{X}_{\ell-1}, \quad \widehat{\mathbf{W}}_\ell := \mathbf{P}_\ell \mathbf{K}_\ell^{-1} \mathbf{B}_\ell. \quad (287)$$

Applying the Lagrangian construction to the three constraints in Eq. (286) gives three adjoint variables: the optical adjoint field $\mathbf{a}_\ell$, the pre-activation error δ_ℓ , and the output error γ_ℓ . Here we pair each residual bilinearly, through $2\operatorname{Re}[\mathbf{a}^T(\cdot)]$ rather than through the Hermitian $2\operatorname{Re}\langle \mathbf{a}, \cdot \rangle_{\mathbb{C}}$ used above. The two differ only by conjugating the multiplier. The bilinear choice produces transposed rather than Hermitian-adjoint operators, as realized by a reciprocal device. It is the layered counterpart of passing to $\mathbf{c} = \bar{\mathbf{a}}$ in Eq. (279). The backward equations are

$$\mathbf{K}_\ell^T \mathbf{a}_\ell = \mathbf{P}_\ell^T \delta_\ell, \quad (288)$$

$$\delta_\ell = (f_\ell)^T_{\mathbf{Z}} \gamma_\ell + \overline{(f_\ell)^T_{\overline{\mathbf{Z}}}}^T \gamma_\ell, \quad (289)$$

$$\gamma_{\ell-1} = \mathbf{B}_\ell^T \mathbf{a}_\ell, \quad \gamma_L = -\psi_{\mathbf{X}_L}. \quad (290)$$

Applied from $\ell = L$ down to 1, these equations implement *physical backpropagation* [23, 39]: begin at the output error $\gamma_L = -\psi_{\mathbf{X}_L}$, pull it through the activation (289) to obtain δ_ℓ , solve the adjoint field equation (288) for $\mathbf{a}_\ell$, and read off the preceding layer's output error $\gamma_{\ell-1} = \mathbf{B}_\ell^T \mathbf{a}_\ell$.

If the optical layer is reciprocal, $\mathbf{K}_\ell^T = \mathbf{K}_\ell$, then Eq. (288) is a physical solve in the same device, excited from the output side with source $\mathbf{P}_\ell^T \delta_\ell$. Combining Eqs. (288) and (290) gives

$$\gamma_{\ell-1} = \mathbf{B}_\ell^T \mathbf{K}_\ell^{-T} \mathbf{P}_\ell^T \delta_\ell = \widehat{\mathbf{W}}_\ell^T \delta_\ell. \quad (291)$$

Thus the physical adjoint solve implements the transpose of the forward optical map.

The parameter gradient follows from the explicit parameter dependence of the layer residuals:

$$\frac{d\chi}{dp_m} = 2 \operatorname{Re} \sum_{\ell=1}^L [\mathbf{a}_\ell^T ((\mathbf{K}_\ell)_{p_m} \mathbf{u}_\ell - (\mathbf{B}_\ell)_{p_m} \mathbf{X}_{\ell-1}) - \delta_\ell^T (\mathbf{P}_\ell)_{p_m} \mathbf{u}_\ell - \gamma_\ell^T (f_\ell)_{p_m}]. \quad (292)$$

When the trainable parameters enter only through the optical operators $\mathbf{K}_\ell$, this reduces to the local field overlap

$$\frac{d\chi}{dp_m} = 2 \operatorname{Re} \sum_{\ell=1}^L \mathbf{a}_\ell^T (\mathbf{K}_\ell)_{p_m} \mathbf{u}_\ell. \quad (293)$$

For a local permittivity perturbation in a phase shifter region, $(\mathbf{K}_\ell)_{\epsilon_m}$ is supported only on that region, so Eq. (293) is measured by a local product of the forward and adjoint optical fields.

I.5 Equivalence with scattering-matrix learning rules

The learning rules derived for linear wave-scattering networks by Wanjura and Marquardt [36] express each parameter gradient as a product of two measured scattering responses. We show that these two responses are the forward and adjoint fields of Appendix I.2, so that their rule is the adjoint identity evaluated at individual ports.

Their neuron modes obey the coupled-mode and input–output equations

$$\mathbf{K}\mathbf{u} = -i\mathbf{\Gamma}\mathbf{q}, \quad \mathbf{q}^{\text{out}} = \mathbf{S}\mathbf{q}, \quad \mathbf{S} = \mathbf{I} + \mathbf{\Gamma}\mathcal{G}\mathbf{\Gamma}, \quad (294)$$

with $\mathbf{q}$ the probe amplitudes, $\mathbf{\Gamma} = \operatorname{diag}(\sqrt{\kappa_1}, \dots, \sqrt{\kappa_N})$ the couplings to the probe waveguides, $\mathbf{K} = \omega\mathbf{I} - \mathbf{H}$ and $\mathcal{G} = -i\mathbf{K}^{-1}$. The dynamical matrix $\mathbf{H}$ collects the detunings and couplings together with the decay rates $-\frac{i}{2}\mathbf{\Gamma}^2$, so $\mathbf{K}$ is symmetric without being Hermitian: the network is reciprocal, $\mathbf{K}^T = \mathbf{K}$ and $\mathcal{G}^T = \mathcal{G}$, and the conjugated adjoint equation (279) applies.

Both fields are probe responses. Probing port p with unit amplitude, $\mathbf{q} = \mathbf{e}_p$, the forward field is the p -th column of $\mathcal{G}$,

$$\mathbf{u} = \sqrt{\kappa_p} \mathcal{G} \mathbf{e}_p. \quad (295)$$

The network output is a quadrature of the scattering amplitude recorded at port r , $y = \operatorname{Re}(e^{i\phi} S_{r,p})$, and $S_{r,p} = \delta_{r,p} + \sqrt{\kappa_r} u_r$, so with the real weight $\lambda := \partial\chi/\partial y$ the loss derivative entering (279) is $\psi_{\mathbf{u}} = \frac{\lambda}{2} e^{i\phi} \sqrt{\kappa_r} \mathbf{e}_r$. Reciprocity then delivers the adjoint field as

$$\mathbf{c} = -\mathbf{K}^{-1} \psi_{\mathbf{u}} = -\frac{i\lambda}{2} e^{i\phi} \sqrt{\kappa_r} \mathcal{G} \mathbf{e}_r, \quad (296)$$

the r -th column of the same $\mathcal{G}$. The adjoint field is the device's response to a probe at the readout port, so both factors of the gradient are scattering measurements on the same hardware.

The gradient is their product rule. Let θ_m be a trainable parameter of $\mathbf{H}$, so that $\mathbf{K}_{\theta_m} = -\mathbf{H}_{\theta_m}$. Substituting (295) and (296) into the gradient (280) and using $\mathcal{G}^T = \mathcal{G}$,

$$\frac{d\chi}{d\theta_m} = 2 \operatorname{Re} [\mathbf{c}^T \mathbf{K}_{\theta_m} \mathbf{u}] = \lambda \operatorname{Re} \left[e^{i\phi} i \sqrt{\kappa_p \kappa_r} (\mathcal{G} \mathbf{H}_{\theta_m} \mathcal{G})_{r,p} \right]. \quad (297)$$

The bracket is the sensitivity of the scattering matrix itself: differentiating $\mathbf{S} = \mathbf{I} + \mathbf{\Gamma}\mathcal{G}\mathbf{\Gamma}$ with $\partial_{\theta_m} \mathbf{K}^{-1} = \mathbf{K}^{-1} \mathbf{H}_{\theta_m} \mathbf{K}^{-1}$ gives

$$\frac{\partial S_{r,p}}{\partial \theta_m} = i \sqrt{\kappa_p \kappa_r} (\mathcal{G} \mathbf{H}_{\theta_m} \mathcal{G})_{r,p}, \quad (298)$$

so (297) is the chain rule $d\chi/d\theta_m = (\partial\chi/\partial y)(\partial y/\partial\theta_m)$, with contributions from several outputs summed.

Read at single ports. Inverting the input–output relation gives $\mathcal{G}_{m,n} = (S_{m,n} - \delta_{m,n})/\sqrt{\kappa_m \kappa_n}$, so the overlap in (298) is a product of measured amplitudes. Writing $\mathbf{E}_{j,\ell}$ for the matrix whose only nonzero entry is a unit entry at (j, ℓ) , an on-site detuning has $\mathbf{H}_{\Delta_j} = \mathbf{E}_{j,j}$ and

$$\frac{\partial S_{r,p}}{\partial \Delta_j} = \frac{i}{\kappa_j} (S_{r,j} - \delta_{r,j})(S_{j,p} - \delta_{j,p}), \quad (299)$$

the response from the probe port to the perturbed site times the response from that site to the readout port, while a reciprocal coupling has $\mathbf{H}_{J_{j,\ell}} = \mathbf{E}_{j,\ell} + \mathbf{E}_{\ell,j}$ and

$$\frac{\partial S_{r,p}}{\partial J_{j,\ell}} = \frac{i}{\sqrt{\kappa_j \kappa_\ell}} [(S_{r,j} - \delta_{r,j})(S_{\ell,p} - \delta_{\ell,p}) + (S_{r,\ell} - \delta_{r,\ell})(S_{j,p} - \delta_{j,p})], \quad (300)$$

the sum over the two paths through the coupled pair. These are the learning rules of Ref. [36], written here in terms of the measured amplitudes $S_{m,n} - \delta_{m,n}$. Their rule is therefore an instance of the adjoint method, in which the drive and the measurement each address a single port and the backward pass is a second scattering experiment on the same device.

I.6 Tangent-space adjoint for quantum equilibrium states

Quantum Equilibrium Propagation (QEP) obtains a loss gradient from the response of ground-state observables to a Hamiltonian perturbation. The derivation below identifies this response with the constrained stationary adjoint and relates its tangent reciprocity to the Onsager symmetry used in Ref. [31].

I.6.1 State equation and readout

Consider a quantum system with Hilbert space $\mathbb{C}^N$ and Hamiltonian

$$\mathbf{H}(\mathbf{p}) = \sum_j p_j \mathbf{A}_j, \quad p_j \in \mathbb{R}, \quad \overline{\mathbf{A}_j}^T = \mathbf{A}_j, \quad (301)$$

where the operators $\mathbf{A}_j$ are independent of the trainable parameters. Inputs may enter through additional parameters held fixed during a training step; these are suppressed here. Assume that $\mathbf{H}$ has a nondegenerate normalized ground state $\mathbf{u}^*$ with energy E_0 and a positive gap $\Delta = E_1 - E_0$ to the first excited state. We use $\langle \mathbf{x}, \mathbf{z} \rangle_{\mathbb{C}} = \overline{\mathbf{x}}^T \mathbf{z}$.

The ground state satisfies the stationary equation

$$\mathbf{F}[\mathbf{u}, \overline{\mathbf{u}}, \mathbf{p}] = \mathbf{H}(\mathbf{p})\mathbf{u} - E(\mathbf{u}, \overline{\mathbf{u}}, \mathbf{p})\mathbf{u} = \mathbf{0}, \quad E = \overline{\mathbf{u}}^T \mathbf{H}(\mathbf{p})\mathbf{u}, \quad \|\mathbf{u}\| = 1. \quad (302)$$

On normalized states, E is the Rayleigh quotient. Its dependence on the state makes $\mathbf{F}$ nonlinear even though $\mathbf{H}$ is a linear operator. Ground-state preparation selects the solution used for the computation.

The readout consists of expectation values of fixed Hermitian observables $\mathbf{O}_\ell$:

$$y_\ell(\mathbf{u}, \overline{\mathbf{u}}) = \langle \mathbf{u}, \mathbf{O}_\ell \mathbf{u} \rangle_{\mathbb{C}} \in \mathbb{R}, \quad \chi = \psi(\mathbf{y}), \quad \overline{\mathbf{O}_\ell}^T = \mathbf{O}_\ell. \quad (303)$$

The real, differentiable loss ψ has no explicit parameter dependence. For a rank-one projector onto a measured degree of freedom, y_ℓ is the probability of that measurement outcome. In the first step of the training, we evaluate the real-valued error signal $\psi_{y_\ell} := \partial\psi/\partial y_\ell$ at the free-state readout $\mathbf{y}(\mathbf{u}^*, \overline{\mathbf{u}^*})$. The chain rule gives the adjoint source at the free state:

$$\mathbf{f}_{\text{adj}} = \chi_{\overline{\mathbf{u}}} = \sum_\ell \psi_{y_\ell} \mathbf{O}_\ell \mathbf{u}^* = \mathbf{B} \mathbf{u}^*, \quad \mathbf{B} := \sum_\ell \psi_{y_\ell} \mathbf{O}_\ell. \quad (304)$$

Thus the source is the action of a Hermitian operator on the state. Its realization as a Hamiltonian perturbation requires the system to implement the operators in $\mathbf{B}$.

I.6.2 Projected stationary adjoint

We introduce the real Lagrangian

$$\mathcal{L} = \chi + \bar{\mathbf{a}}^T \mathbf{F} + \mathbf{a}^T \bar{\mathbf{F}}. \quad (305)$$

To differentiate, we extend the polynomial expression for $\mathbf{F}$ to independent ambient Wirtinger arguments, then restrict physical variations to the normalized state with its arbitrary overall phase chosen consistently for nearby states.

Normalization imposes $\text{Re}\langle \mathbf{u}^*, \delta \mathbf{u} \rangle_{\mathbb{C}} = 0$. Choosing the phase locally so that $\text{Im}\langle \mathbf{u}^*, \delta \mathbf{u} \rangle_{\mathbb{C}} = 0$ removes the unobservable phase, leaving

$$\mathcal{T} = \{\delta \mathbf{u} \in \mathbb{C}^N : \langle \mathbf{u}^*, \delta \mathbf{u} \rangle_{\mathbb{C}} = 0\}, \quad \mathbb{P}_{\mathcal{T}} = \mathbf{I} - \mathbf{u}^* \mathbf{u}^{*T}. \quad (306)$$

Consequently, the state and conjugate variations obey

$$\delta \mathbf{u} = \mathbb{P}_{\mathcal{T}} \delta \mathbf{u}, \quad \delta \bar{\mathbf{u}} = \overline{\delta \mathbf{u}} = \overline{\mathbb{P}_{\mathcal{T}} \delta \mathbf{u}} = \mathbb{P}_{\mathcal{T}}^T \delta \bar{\mathbf{u}}, \quad (307)$$

where $\overline{\mathbb{P}_{\mathcal{T}}}^T = \mathbb{P}_{\mathcal{T}}$ and $\overline{\mathbb{P}_{\mathcal{T}}} = \mathbb{P}_{\mathcal{T}}^T$.

Using the source (304), the constrained Wirtinger variation is

$$\begin{aligned} \delta_{\bar{\mathbf{u}}} \mathcal{L} &= \left(\mathbf{f}_{\text{adj}} + \overline{\mathbf{F}_{\mathbf{u}}}^T \mathbf{a} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{a}} \right)^T \mathbb{P}_{\mathcal{T}}^T \delta \bar{\mathbf{u}} \\ &= \left[\mathbb{P}_{\mathcal{T}} \left(\mathbf{f}_{\text{adj}} + \overline{\mathbf{F}_{\mathbf{u}}}^T \mathbf{a} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{a}} \right) \right]^T \delta \bar{\mathbf{u}}. \end{aligned} \quad (308)$$

All Jacobians here and below are evaluated at the free state unless otherwise specified. Requiring stationarity for every admissible variation gives

$$\mathbb{P}_{\mathcal{T}} \left(\overline{\mathbf{F}_{\mathbf{u}}}^T \mathbf{a} + \mathbf{F}_{\bar{\mathbf{u}}}^T \bar{\mathbf{a}} + \mathbf{f}_{\text{adj}} \right) = \mathbf{0}. \quad (309)$$

Because $\mathcal{L}$ is real, the conjugate stationarity condition follows as well. The adjoint is chosen in $\mathcal{T}$, so $\mathbf{a} = \mathbb{P}_{\mathcal{T}} \mathbf{a}$: the component along $\mathbf{u}^*$ does not contribute to its pairing with the tangent residual variation, as the Jacobians below verify. This removes the multiplier redundancy; uniqueness also requires invertibility on $\mathcal{T}$. Inserting this restriction and its conjugate gives the equivalent form

$$\overline{\mathbb{P}_{\mathcal{T}} \mathbf{F}_{\mathbf{u}} \mathbb{P}_{\mathcal{T}}}^T \mathbf{a} + (\mathbb{P}_{\mathcal{T}} \mathbf{F}_{\bar{\mathbf{u}}} \mathbb{P}_{\mathcal{T}}^T)^T \bar{\mathbf{a}} + \mathbb{P}_{\mathcal{T}} \mathbf{f}_{\text{adj}} = \mathbf{0}. \quad (310)$$

I.6.3 Projected Wirtinger Jacobians

At fixed parameters, the independent variations of $E = \bar{\mathbf{u}}^T \mathbf{H} \mathbf{u}$ are

$$\delta_{\mathbf{u}} E = \bar{\mathbf{u}}^T \mathbf{H} \delta \mathbf{u}, \quad \delta_{\bar{\mathbf{u}}} E = (\mathbf{H} \mathbf{u})^T \delta \bar{\mathbf{u}}. \quad (311)$$

The product rule gives the ambient Jacobians

$$\mathbf{F}_{\mathbf{u}} = \mathbf{H} - E \mathbf{I} - \mathbf{u} \bar{\mathbf{u}}^T \mathbf{H}, \quad \mathbf{F}_{\bar{\mathbf{u}}} = -\mathbf{u} (\mathbf{H} \mathbf{u})^T. \quad (312)$$

Since $\mathbf{H} \mathbf{u}^* = E \mathbf{u}^*$, Hermiticity also gives $\overline{\mathbf{u}^*}^T \mathbf{H} = E \overline{\mathbf{u}^*}^T$, so

$$\mathbf{F}_{\mathbf{u}} = \mathbf{H} - E \mathbf{I} - E \mathbf{u}^* \overline{\mathbf{u}^*}^T, \quad \mathbf{F}_{\bar{\mathbf{u}}} = -E \mathbf{u}^* \mathbf{u}^{*T}. \quad (313)$$

The constrained linearization projects the output onto $\mathcal{T}$ and restricts the two inputs to $\mathcal{T}$ and $\bar{\mathcal{T}}$, respectively. Since $\mathbb{P}_{\mathcal{T}} \mathbf{u}^* = 0$, both rank-one terms disappear:

$$\begin{aligned} \mathbb{P}_{\mathcal{T}} \mathbf{F}_{\mathbf{u}} \mathbb{P}_{\mathcal{T}} &= \mathbb{P}_{\mathcal{T}} (\mathbf{H} - E \mathbf{I}) \mathbb{P}_{\mathcal{T}}, \\ \mathbb{P}_{\mathcal{T}} \mathbf{F}_{\bar{\mathbf{u}}} \mathbb{P}_{\mathcal{T}}^T &= \mathbf{0}. \end{aligned} \quad (314)$$

The first projected Jacobian is Hermitian, because $\mathbf{H}$ and $\mathbb{P}_{\mathcal{T}}$ are Hermitian and $E \in \mathbb{R}$; the second is the zero matrix, which is symmetric. Explicitly,

$$\begin{aligned}\overline{\mathbb{P}_{\mathcal{T}} \mathbf{F}_{\mathbf{u}} \mathbb{P}_{\mathcal{T}}}^T &= \mathbb{P}_{\mathcal{T}} \mathbf{F}_{\mathbf{u}} \mathbb{P}_{\mathcal{T}}, \\ (\mathbb{P}_{\mathcal{T}} \mathbf{F}_{\mathbf{u}} \mathbb{P}_{\mathcal{T}}^T)^T &= \mathbb{P}_{\mathcal{T}} \mathbf{F}_{\mathbf{u}} \mathbb{P}_{\mathcal{T}}^T.\end{aligned}\tag{315}$$

These are the tangent reciprocity conditions (64) of Theorem 8.1, restricted to the physical tangent space. Substituting into (310) yields

$$\mathbb{P}_{\mathcal{T}}(\mathbf{H} - E\mathbf{I})\mathbf{a} + \mathbb{P}_{\mathcal{T}}\mathbf{f}_{\text{adj}} = \mathbf{0}, \quad \mathbf{a} \in \mathcal{T}.\tag{316}$$

I.6.4 Spectral-gap invertibility

Let $\{\mathbf{u}_n\}_{n=0}^{N-1}$ be an orthonormal eigenbasis, ordered by energy:

$$\mathbf{H}\mathbf{u}_n = E_n\mathbf{u}_n, \quad \mathbf{u}_0 = \mathbf{u}^*, \quad E_0 = E.\tag{317}$$

For a nondegenerate ground state, $\mathcal{T}$ is spanned by the excited eigenvectors. Every $\delta\mathbf{u} \in \mathcal{T}$ therefore expands as $\delta\mathbf{u} = \sum_{n \geq 1} c_n \mathbf{u}_n$, and

$$\mathbb{P}_{\mathcal{T}}(\mathbf{H} - E\mathbf{I})\mathbb{P}_{\mathcal{T}}\delta\mathbf{u} = \sum_{n \geq 1} (E_n - E_0)c_n \mathbf{u}_n.\tag{318}$$

In particular, $\mathbf{H} - E\mathbf{I}$ preserves $\mathcal{T}$, so its restriction equals that of the projected operator. With $\Delta := E_1 - E_0 > 0$, its spectrum and inverse norm are

$$\text{spec}((\mathbf{H} - E\mathbf{I})|_{\mathcal{T}}) = \{E_n - E_0\}_{n \geq 1}, \quad \left\|((\mathbf{H} - E\mathbf{I})|_{\mathcal{T}})^{-1}\right\| = \Delta^{-1}.\tag{319}$$

The inverse multiplies each $\mathbf{u}_n$ by $(E_n - E_0)^{-1}$, and its induced Hilbert space norm is the largest of these positive inverse eigenvalues. The adjoint is then unique and can be written as

$$\mathbf{a} = - \sum_{n \geq 1} \frac{\langle \mathbf{u}_n, \mathbf{f}_{\text{adj}} \rangle_{\mathbb{C}}}{E_n - E_0} \mathbf{u}_n, \quad \|\mathbf{a}\| \leq \Delta^{-1} \|\mathbb{P}_{\mathcal{T}}\mathbf{f}_{\text{adj}}\|.\tag{320}$$

The gap controls the inverse norm; the condition number of the restricted operator is $(E_{N-1} - E_0)/\Delta$. A physical implementation also requires ground-state preparation for the free and nudged Hamiltonians. The spectral gap of $\mathbf{H}$ does not by itself establish the exponential stability of that preparation dynamics on physical states.

I.6.5 The Hamiltonian nudge generates the adjoint

With the error signal components fixed at their free-state values, the nudged Hamiltonian of Eq. (75) is

$$\mathbf{H}_{\epsilon} = \mathbf{H} + \epsilon \mathbf{B} = \mathbf{H} + \epsilon \sum_{\ell} \psi_{y_{\ell}} \mathbf{O}_{\ell}.$$

Let $\mathbf{u}^{*,\epsilon}$ be its normalized ground state and write $\langle \mathbf{C} \rangle_{\epsilon} := \langle \mathbf{u}^{*,\epsilon}, \mathbf{C} \mathbf{u}^{*,\epsilon} \rangle_{\mathbb{C}}$ for an operator $\mathbf{C}$. The nudged eigenvalue equation reads

$$(\mathbf{H} + \epsilon \mathbf{B} - E_{\epsilon} \mathbf{I}) \mathbf{u}^{*,\epsilon} = \mathbf{0}, \quad E_{\epsilon} = \langle \mathbf{H}_{\epsilon} \rangle_{\epsilon}.\tag{321}$$

Since $E_{\epsilon} = \langle \mathbf{H} \rangle_{\epsilon} + \epsilon \langle \mathbf{B} \rangle_{\epsilon}$, this is equivalently

$$\mathbf{F}[\mathbf{u}^{*,\epsilon}, \overline{\mathbf{u}^{*,\epsilon}}, \mathbf{p}] + \epsilon(\mathbf{B} - \langle \mathbf{B} \rangle_{\epsilon} \mathbf{I}) \mathbf{u}^{*,\epsilon} = \mathbf{0}.\tag{322}$$

The expectation value subtraction accounts for the change in the ground state energy and projects the nudge onto the normalized state's tangent space.

The nondegenerate ground state and positive gap ensure a smooth branch for sufficiently small ϵ . We choose its phase smoothly so that $\langle \mathbf{u}^*, \mathbf{u}^{*,\epsilon} \rangle_{\mathbb{C}}$ is real and positive. This is possible locally because the overlap equals one at $\epsilon = 0$. Differentiating the normalization gives $\text{Re}\langle \mathbf{u}^*, \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} \rangle_{\mathbb{C}} = 0$, and the phase choice gives $\text{Im}\langle \mathbf{u}^*, \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} \rangle_{\mathbb{C}} = 0$, where

$$\mathbf{u}^{*,\epsilon} = \mathbf{u}^* + \epsilon \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} + O(\epsilon^2), \quad \mathbf{v} = \left. \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} \right|_{\epsilon=0} \in \mathcal{T}. \quad (323)$$

Subtracting the free stationary equation from Eq. (322) and retaining first-order terms yields

$$\mathbf{F}_{\mathbf{u}} \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} + \mathbf{F}_{\bar{\mathbf{u}}} \overline{\frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon}} + (\mathbf{B} - \langle \mathbf{B} \rangle_0 \mathbf{I}) \mathbf{u}^* = \mathbf{0}. \quad (324)$$

Projecting onto $\mathcal{T}$ and using Eq. (314) gives

$$\mathbb{P}_{\mathcal{T}}(\mathbf{H} - E_0 \mathbf{I}) \mathbb{P}_{\mathcal{T}} \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} + \mathbb{P}_{\mathcal{T}} \mathbf{f}_{\text{adj}} = \mathbf{0}. \quad (325)$$

This is the reduced adjoint equation (316). Invertibility on $\mathcal{T}$ therefore gives

$$\mathbf{a} = \left. \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} \right|_{\epsilon=0} = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{u}^{*,\epsilon} - \mathbf{u}^*}{\epsilon}. \quad (326)$$

The plus sign in the Hamiltonian perturbation agrees with the adjoint-source sign in the stationary Lagrangian.

I.6.6 The gradient overlap is an observable response

At the constrained stationary state, the adjoint cancels the state variation of $\mathcal{L}$, so $d\chi/dp_j = \partial\mathcal{L}/\partial p_j$. The loss has no explicit parameter dependence, and differentiating the residual at fixed state gives

$$\mathbf{F}_{p_j} = (\mathbf{A}_j - \langle \mathbf{A}_j \rangle_0 \mathbf{I}) \mathbf{u}^*. \quad (327)$$

Hence, the stationary adjoint gradient is

$$\frac{d\chi}{dp_j} = 2 \text{Re} \langle \mathbf{a}, (\mathbf{A}_j - \langle \mathbf{A}_j \rangle_0 \mathbf{I}) \mathbf{u}^* \rangle_{\mathbb{C}} = 2 \text{Re} \langle \mathbf{a}, \mathbf{A}_j \mathbf{u}^* \rangle_{\mathbb{C}}. \quad (328)$$

The second equality uses $\langle \mathbf{a}, \mathbf{u}^* \rangle_{\mathbb{C}} = 0$. Hermiticity of $\mathbf{A}_j$ and Eq. (326) give

$$\begin{aligned} \left. \frac{d\langle \mathbf{A}_j \rangle_{\epsilon}}{d\epsilon} \right|_{\epsilon=0} &= \left\langle \left. \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} \right|_{\epsilon=0}, \mathbf{A}_j \mathbf{u}^* \right\rangle_{\mathbb{C}} + \left\langle \mathbf{u}^*, \mathbf{A}_j \left. \frac{d\mathbf{u}^{*,\epsilon}}{d\epsilon} \right|_{\epsilon=0} \right\rangle_{\mathbb{C}} \\ &= 2 \text{Re} \langle \mathbf{a}, \mathbf{A}_j \mathbf{u}^* \rangle_{\mathbb{C}} = \frac{d\chi}{dp_j}. \end{aligned} \quad (329)$$

This proves the observable gradient formula (76) and the QEP rule of Ref. [31]:

$$\frac{d\chi}{dp_j} = \left. \frac{d\langle \mathbf{A}_j \rangle_{\epsilon}}{d\epsilon} \right|_{\epsilon=0} = \frac{\langle \mathbf{A}_j \rangle_{\epsilon} - \langle \mathbf{A}_j \rangle_0}{\epsilon} + O(\epsilon). \quad (330)$$

The experiment measures the change in $\langle \mathbf{A}_j \rangle$, so it does not require phase-resolved measurements of either the state or the adjoint. The gradient readout is local whenever $\mathbf{A}_j$ is local: a two-site coupling requires the correlator of the two sites it couples.

I.6.7 Onsager reciprocity

The static susceptibility symmetry was derived in the Supplementary Information of Ref. [31]. Here, we recast this derivation in terms of the two projected Wirtinger conditions. Since $\mathbb{P}_{\mathcal{T}}\mathbf{F}_{\mathbf{u}}\mathbb{P}_{\mathcal{T}}^T = 0$, the tangent response has no conjugate-variation term and is governed solely by $\mathbb{P}_{\mathcal{T}}\mathbf{F}_{\mathbf{u}}\mathbb{P}_{\mathcal{T}}$. We first define the reduced resolvent operator:

$$G := -[(\mathbb{P}_{\mathcal{T}}\mathbf{F}_{\mathbf{u}}\mathbb{P}_{\mathcal{T}})|_{\mathcal{T}}]^{-1} = [(E_0\mathbf{I} - \mathbf{H})|_{\mathcal{T}}]^{-1} = \overline{G}^T, \quad (331)$$

where the existence of the inverse follows separately from the positive gap (319) and Hermiticity follows from that of the tangent operator $\mathbb{P}_{\mathcal{T}}\mathbf{F}_{\mathbf{u}}\mathbb{P}_{\mathcal{T}}$. The Onsager reciprocity condition used in Ref. [31] is

$$\frac{\partial\langle\mathbf{A}_{\ell}\rangle}{\partial p_j} = \frac{\partial\langle\mathbf{A}_j\rangle}{\partial p_{\ell}}, \quad (332)$$

where the expectation values are evaluated in equilibrium. To derive this reciprocity from the projected tangent conditions (64), we choose the phase locally so that $\partial\mathbf{u}^*/\partial p_j \in \mathcal{T}$, and differentiate the ground-state equation $\mathbf{H}\mathbf{u}^* = E_0\mathbf{u}^*$ with respect to p_j . Since $\partial\mathbf{H}/\partial p_j = \mathbf{A}_j$, this gives

$$\mathbf{A}_j\mathbf{u}^* + \mathbf{H}\frac{\partial\mathbf{u}^*}{\partial p_j} = \frac{\partial E_0}{\partial p_j}\mathbf{u}^* + E_0\frac{\partial\mathbf{u}^*}{\partial p_j}. \quad (333)$$

Projecting onto $\mathcal{T}$ removes the energy derivative term because $\mathbb{P}_{\mathcal{T}}\mathbf{u}^* = 0$. Using $\mathbb{P}_{\mathcal{T}}\partial\mathbf{u}^*/\partial p_j = \partial\mathbf{u}^*/\partial p_j$, we obtain

$$\mathbb{P}_{\mathcal{T}}(E_0\mathbf{I} - \mathbf{H})\mathbb{P}_{\mathcal{T}}\frac{\partial\mathbf{u}^*}{\partial p_j} = \mathbb{P}_{\mathcal{T}}\mathbf{A}_j\mathbf{u}^*. \quad (334)$$

On $\mathcal{T}$, the operator on the left has the inverse G defined in (331). Applying this inverse therefore yields

$$\frac{\partial\mathbf{u}^*}{\partial p_j} = G\mathbb{P}_{\mathcal{T}}\mathbf{A}_j\mathbf{u}^*. \quad (335)$$

For the parameter-independent Hermitian operators $\mathbf{A}_j$, differentiating $\langle\mathbf{A}_j\rangle_0 = \langle\mathbf{u}^*, \mathbf{A}_j\mathbf{u}^*\rangle_{\mathbb{C}}$ and using (335) gives

$$\begin{aligned} \frac{\partial\langle\mathbf{A}_j\rangle_0}{\partial p_k} &= \left\langle \frac{\partial\mathbf{u}^*}{\partial p_k}, \mathbf{A}_j\mathbf{u}^* \right\rangle_{\mathbb{C}} + \left\langle \mathbf{u}^*, \mathbf{A}_j \frac{\partial\mathbf{u}^*}{\partial p_k} \right\rangle_{\mathbb{C}} \\ &= 2 \operatorname{Re} \left\langle \mathbf{A}_j\mathbf{u}^*, \frac{\partial\mathbf{u}^*}{\partial p_k} \right\rangle_{\mathbb{C}} \\ &= 2 \operatorname{Re} \langle \mathbb{P}_{\mathcal{T}}\mathbf{A}_j\mathbf{u}^*, G\mathbb{P}_{\mathcal{T}}\mathbf{A}_k\mathbf{u}^* \rangle_{\mathbb{C}}. \end{aligned} \quad (336)$$

The projector in the first argument may be inserted because the state derivative lies in $\mathcal{T}$. Hermiticity of G on this space then allows it to be moved between the two arguments. Taking the real part and using conjugate symmetry of the inner product yields

$$\begin{aligned} \frac{\partial\langle\mathbf{A}_j\rangle_0}{\partial p_k} &= 2 \operatorname{Re} \langle G\mathbb{P}_{\mathcal{T}}\mathbf{A}_j\mathbf{u}^*, \mathbb{P}_{\mathcal{T}}\mathbf{A}_k\mathbf{u}^* \rangle_{\mathbb{C}} \\ &= 2 \operatorname{Re} \langle \mathbb{P}_{\mathcal{T}}\mathbf{A}_k\mathbf{u}^*, G\mathbb{P}_{\mathcal{T}}\mathbf{A}_j\mathbf{u}^* \rangle_{\mathbb{C}} \\ &= \frac{\partial\langle\mathbf{A}_k\rangle_0}{\partial p_j}. \end{aligned} \quad (337)$$

I.6.8 Finite nudges and measurement noise

A sufficient condition for staying in the perturbative regime is $|\epsilon|\|\mathbf{B}\| \ll \Delta$, where the norm is the induced Hilbert space operator norm. The free and nudged preparations must follow the same smooth ground-state

branch. The inverse-norm bound (319) allows a larger response as the gap closes, so smaller nudges may be needed near a phase boundary. Centered nudging reduces the finite-difference bias:

$$\frac{d\chi}{dp_j} = \frac{\langle \mathbf{A}_j \rangle_\epsilon - \langle \mathbf{A}_j \rangle_{-\epsilon}}{2\epsilon} + O(\epsilon^2). \quad (338)$$

Both nudges use the same error signal from the free state. This requires two nudged preparation settings in addition to the free preparation used to determine those coefficients.

For fixed error signal, independent estimates of each expectation value from M measurements with single shot variance of order σ^2 give a gradient estimator variance of order $\sigma^2/(\epsilon^2 M)$. Decreasing $|\epsilon|$ reduces the finite-nudge bias but increases this variance.

J How to test a gradient: the Taylor test

The *Taylor test* provides a simple, model-agnostic check of an adjoint implementation and is widely used in PDE-constrained optimization [111, 112].

Let $\mathbf{g} := d\chi/d\mathbf{p}$ denote the gradient produced by the adjoint procedure, and let $\mathbf{v} \in \mathbb{R}^M$ be a random direction (e.g. a Gaussian vector). For a small step size $h > 0$, define the zeroth- and first-order remainders

$$R_0(h) := |\chi(\mathbf{p} + h\mathbf{v}) - \chi(\mathbf{p})|, \quad (339)$$

$$R_1(h) := |\chi(\mathbf{p} + h\mathbf{v}) - \chi(\mathbf{p}) - h\langle \mathbf{g}, \mathbf{v} \rangle|. \quad (340)$$

A Taylor expansion of χ about $\mathbf{p}$ gives

$$R_0(h) = h|\langle \mathbf{g}, \mathbf{v} \rangle| + \mathcal{O}(h^2), \quad R_1(h) = \mathcal{O}(h^2). \quad (341)$$

The diagnostic is a log–log plot of $R_0(h)$ and $R_1(h)$ as h is swept through several decades:

- R_0 should decrease with slope 1 in $\log h$.
- R_1 should decrease with slope 2, *provided* $\mathbf{g}$ is the true gradient.
- If $\mathbf{g}$ is wrong, R_1 instead decreases with slope 1, and the curve typically transitions from a slope of 1 at large h to a noise-dominated floor at small h , producing the characteristic “hockey-stick” shape.

For very small h , R_1 becomes dominated by floating-point round-off in the forward evaluation. The Taylor test should therefore be assessed using the slope over the regime where R_1 still decreases monotonically.

Figure 4 shows the test for the nonlinear Schrödinger example in the accompanying notebooks. Here $R_0(h)$ decreases with slope 1 as required by Eq. (341). For the correct, physically obtained gradient, $R_1(h)$ decreases with slope 2 over eight decades in h before saturating at the round-off floor of the forward solve. For a gradient computed with the tangent reciprocity condition (A1) deliberately violated, $R_1(h)$ decreases with only slope 1, remaining parallel to $R_0(h)$ rather than curving down to match the correct-gradient trend. Each accompanying notebook reports this diagnostic for its adjoint implementation.

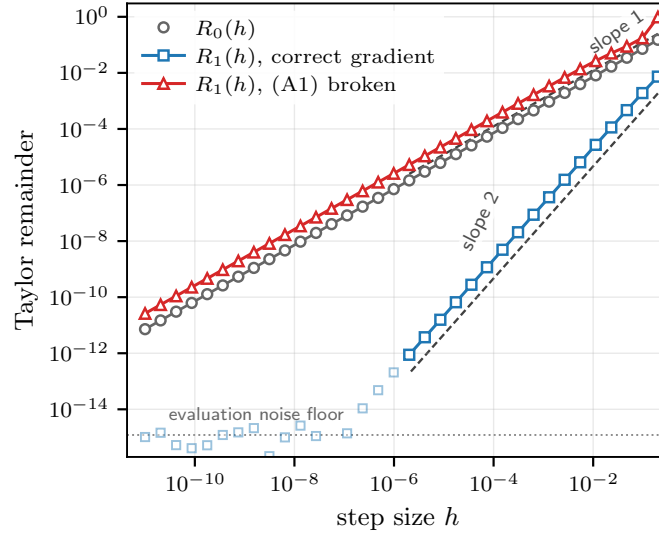


Figure 4: Taylor test for the nonlinear Schrödinger example, following Eqs. (339)–(341). Circles: $R_0(h)$, decreasing with slope 1. Blue squares: $R_1(h)$ for the correct, physically obtained gradient, decreasing with slope 2 until it saturates at the round-off floor of the forward evaluation (faint markers). Red triangles: $R_1(h)$ for a gradient computed with condition (A1) broken, which decreases with only slope 1 and never curves down to the correct gradient trend. Dashed lines are slope-1 and slope-2 references.