A Analytical whole-body reference controller

This appendix gives the full analytical compliance controller summarized in §3. It is used *only* as a reference that generates the per-link virtual targets $\boldsymbol{x}^{\mathrm{c}}_{\ell,d}$ of Eq. (1) used in the policy reward; the reference torque $\boldsymbol{\tau}_{\mathrm{c}}$ and the energy tank below are derivation/reference quantities and are *not* executed at deployment.

A humanoid has configuration $\boldsymbol{q} \in \mathrm{SE}(3) \times \mathbb{R}^{n_j}$, generalized velocity $\boldsymbol{\nu}$, and floating-base dynamics $\boldsymbol{M}\dot{\boldsymbol{\nu}} + \boldsymbol{C}\boldsymbol{\nu} + \boldsymbol{g} = \boldsymbol{S}^\top \boldsymbol{\tau} + \sum_c \boldsymbol{J}_c^\top \boldsymbol{\lambda}_c + \sum_i \boldsymbol{J}_{\ell_i}^\top \boldsymbol{f}_{\mathrm{ext},i}$ with support-contact reaction wrenches $\boldsymbol{\lambda}_c$ and external wrenches $\boldsymbol{f}_{\mathrm{ext},i}$ at links ℓ_i. Classical Cartesian impedance at a single end-effector [1] prescribes $\boldsymbol{M}_d\ddot{\boldsymbol{x}} + \boldsymbol{D}_d\dot{\boldsymbol{x}} + \boldsymbol{K}_d(\boldsymbol{x} - \boldsymbol{x}_d) = \boldsymbol{f}_{\mathrm{ext}}$; we lift this to a hierarchy of frames coupled by balance.

Reference compliant torque. For controlled links $\mathcal{L}$ (hands, elbows, knees, pelvis, torso), support contacts $\mathcal{C}_s$, and centroidal momentum $\boldsymbol{h}$, the reference torque is

$$\boldsymbol{\tau}_{\mathrm{c}} = \underbrace{\boldsymbol{J}_{\boldsymbol{h}}^\top \boldsymbol{w}_{\boldsymbol{h}}^{\mathrm{imp}}}_{\text{centroidal balance}} + \underbrace{\sum_{\ell \in \mathcal{L}} \boldsymbol{N}_{\mathrm{bal}}^\top \boldsymbol{J}_\ell^\top [\boldsymbol{K}_\ell(\boldsymbol{x}_{\ell,d} - \boldsymbol{x}_\ell) + \boldsymbol{D}_\ell(\dot{\boldsymbol{x}}_{\ell,d} - \dot{\boldsymbol{x}}_\ell)]}_{\text{per-link Cartesian impedance}} + \underbrace{\boldsymbol{N}_{\mathrm{task}}^\top [\boldsymbol{K}_q(\boldsymbol{q} - \boldsymbol{q}_{\mathrm{nom}}) + \boldsymbol{D}_q\dot{\boldsymbol{q}}]}_{\text{joint-space posture}} \tag{7}$$

where the centroidal-impedance wrench is $\boldsymbol{w}_{\boldsymbol{h}}^{\mathrm{imp}} = \boldsymbol{K}_h(\boldsymbol{h}_d - \boldsymbol{h}) + \boldsymbol{D}_h(\dot{\boldsymbol{h}}_d - \dot{\boldsymbol{h}}) + m\boldsymbol{g}$, $\boldsymbol{N}_{\mathrm{bal}} = \boldsymbol{I} - \boldsymbol{J}_{\boldsymbol{h}}^\top (\boldsymbol{J}_{\boldsymbol{h}} \boldsymbol{M}^{-1} \boldsymbol{J}_{\boldsymbol{h}}^\top)^{-1} \boldsymbol{J}_{\boldsymbol{h}} \boldsymbol{M}^{-1}$ is the dynamically consistent null-space projector of the balance task [3], and $\boldsymbol{N}_{\mathrm{task}}$ projects orthogonal to both balance and per-link tasks. Each per-link Cartesian wrench is first mapped to joint torque by $\boldsymbol{J}_\ell^\top$ and then projected by $\boldsymbol{N}_{\mathrm{bal}}^\top$, so every term acts in the n_j-dimensional generalized-force space. The desired centroidal wrench is realized by the support contacts via a balancing QP in the spirit of [5]:

$$\min_{\{\boldsymbol{\lambda}_c\}} \left\| \boldsymbol{w}_{\boldsymbol{h}}^{\mathrm{imp}} - \textstyle\sum_c \boldsymbol{G}_c \boldsymbol{\lambda}_c - \sum_i \boldsymbol{G}_{\mathrm{ext},i} \boldsymbol{f}_{\mathrm{ext},i} \right\|^2 + \gamma_{\mathrm{reg}} \textstyle\sum_c \|\boldsymbol{\lambda}_c\|^2 \;\; \text{s.t.} \;\; \boldsymbol{\lambda}_c \in \mathcal{F}_{\mu_c}, \; \mathrm{CoP}_c \in \mathcal{S}_c \tag{8}$$

with grasp maps $\boldsymbol{G}_c$, linearized friction cones $\mathcal{F}_{\mu_c}$, support polygons $\mathcal{S}_c$, and regularizer weight γ_{reg}. Admitting knees into $\mathcal{C}_s$ adds support sites and generalizes the two-footed setting to multi-contact, lower-body-involved postures.

Energy tank (general time-varying-stiffness case). Time-varying $\boldsymbol{K}_\ell(t)$ with null-space projection injects power and breaks passivity [27, 28]. For that general case one may add a scalar tank energy $T(t) \in [T_{\min}, T_{\max}]$,

$$\dot{T} = \textstyle\sum_\ell \dot{\boldsymbol{x}}_\ell^\top \boldsymbol{D}_\ell \dot{\boldsymbol{x}}_\ell - P_{\mathrm{inject}}(\dot{\boldsymbol{K}}_\ell, \boldsymbol{N}_{\mathrm{bal}}, \boldsymbol{N}_{\mathrm{task}}), \tag{9}$$

filled by dissipated power and drained by non-passive control, freezing gains as $T \to T_{\min}$. In our deployed pipeline $\boldsymbol{K}_\ell$ is held fixed and the residual edits only the equilibrium, so this tank is inert at runtime and is retained only as scaffolding for the time-varying-stiffness extension (cf. §F).

B Architecture details

Base policy. 3-layer MLP, hidden width 512, trained with PPO [44]. Observation $\boldsymbol{o}$ comprises joint configurations, velocities, projected gravity, the last action, and the motion command $\boldsymbol{c}$. The latent $\boldsymbol{z}_F$ enters as an additional input.

Force Latent Encoder. The encoder takes a history window $\boldsymbol{h}_t = (\boldsymbol{q}_{t-H:t}, \boldsymbol{\nu}_{t-H:t}, \boldsymbol{\tau}_{t-H:t-1}, \boldsymbol{c}_{t-H:t})$ of $H = 16$ steps (160 ms at 100 Hz), flattens it, and passes it through a 3-layer MLP (widths 512 → 512 → 256, LayerNorm on the last hidden layer) emitting $(\boldsymbol{\mu}_t, \log \boldsymbol{\sigma}_t^2)$ over a latent of dimension $d_F = 32$. A reparameterized sample $\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{\sigma}_t \odot \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, drives the wrench head $\psi : \boldsymbol{z}_F \to \hat{\boldsymbol{f}}_{\mathrm{ext}} \in \mathbb{R}^{3|\mathcal{L}|}$ on all $|\mathcal{L}|$ controlled links and the projection head $g : \boldsymbol{z}_F \to \mathbb{S}^{d_g - 1}$ with $d_g = 64$.

Auxiliary losses.

$$\mathcal{L}_{\text{wrench}} = \big\|\psi(\boldsymbol{z}_t) - \boldsymbol{f}_{\text{ext}}\big\|_2^2, \tag{10}$$

$$\mathcal{L}_{\text{supcon}} = \mathcal{L}^{\text{sup}}\big(g(\boldsymbol{z}_t),\, y_t\big), \tag{11}$$

$$\mathcal{L}_{\text{kl}} = \mathrm{D}_{\text{KL}}\big(\mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{\sigma}_t^2)\,\big\|\,\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})\big), \tag{12}$$

$$\mathcal{L}_{\text{smooth}} = \big\|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}\big\|_2^2. \tag{13}$$

Weights: $(\lambda_w, \lambda_s, \lambda_k, \lambda_m) = (1.0,\ 0.3,\ 10^{-3},\ 0.1)$. $\mathcal{L}^{\text{sup}}$ is the supervised contrastive loss of [21] with perturbation-class labels y_t indicating which limb is perturbed and in which direction.

Residual policy. 3-layer MLP, hidden width 256, trained with PPO against the same reward of Eq. (2). Output is bounded as in Eq. (4) with $\epsilon_x = 5$ cm; the per-link stiffness $\boldsymbol{K}_\ell$ is held fixed by the curriculum throughout Stage 2.

C Force sampling details

Phong-weighted force-origin sampling. The inverse-CDF expressions for the Phong lobe of Eq. (5) are

$$\cos\theta = u^{1/(n+1)}, \qquad \phi \sim \mathcal{U}[0, 2\pi), \qquad r = R_a\, v^{1/3}, \tag{14}$$

with $u, v \sim \mathcal{U}[0, 1]$, θ measured from $-\hat{\boldsymbol{e}}_z$. The cube-root radial warp yields uniform density in volume; with probability ϵ the direction is drawn isotropically instead.

The pelvis stiffness (k, c) is set so the spring force at maximum displacement R_a reaches a fixed fraction of the robot's nominal weight: the stiffness term yields a coherent load that grows as the policy drifts from the anchor and admits a well-defined rest position against the leg stiffness, while the damping term reproduces the velocity-dependent tug of an elastically attached inertia. Per chunk we draw hold time $T_{\text{hold}} \sim \mathcal{U}(0.2, 2.0)$ s, ramp-down $\mathcal{U}(0.25, 1.0)$ s, and upper-body stiffness $K_p \sim \mathcal{U}(5, 250)$ N/m, with upper-body type sampled from {none, single-arm, both arms}; knees and feet are never perturbed directly. The shared downward bias of Eq. (5) makes the dominant carried-load regime coherent without an explicit pelvis-to-upper-body direction-inheritance heuristic.

Force law. Upper-body anchors apply a unilateral spring that auto-ramps as the body approaches and releases [11], while the pelvis is driven by a mass–spring–damper coupling to the sampled anchor to model a heavy object held through a compliant grasp. The two laws share the same anchor-based form and are gated on link type:

$$\boldsymbol{f}_\ell(t) = \begin{cases} k\big(\boldsymbol{a}_{\text{pelvis}}(t) - \boldsymbol{x}_{\text{pelvis}}(t)\big) - c\,\dot{\boldsymbol{x}}_{\text{pelvis}}(t), & \ell = \text{pelvis}, \\ K_p\big(\boldsymbol{a}_\ell - \boldsymbol{x}_\ell(t)\big)\,\mathbf{1}[\boldsymbol{a}_\ell \text{ on active side}], & \ell \in \mathcal{L}_{\text{upper}}. \end{cases} \tag{15}$$

D Algorithm

Reward. The trained objective is the compliance-fidelity reward of Eq. (2), where $\boldsymbol{x}^{\text{c}}_{\ell,d}$ is the analytical controller's impedance-implied trajectory under the realized (ground-truth) perturbation, available in simulation. The term $w_c\|\cdot\|^2$ rewards following this compliant deformation rather than the original motion command—generalizing GentleHumanoid's [11] reference-dynamics reward to the whole body. The encoder's wrench-reconstruction MSE does *not* appear in r_t: it shapes ϕ only through $\mathcal{L}_\phi$ (Eq. (3)), behind the gradient barrier, so the policy reward never confuses "latent fits the wrench" with "robot behaves compliantly."

Hyperparameters. $N_1 = 5{\times}10^8$ environment-steps for Stage 1; $N_2 = 5{\times}10^8$ environment-steps for Stage 2; 8192 parallel environments in MjWarp (MuJoCo-based [50]); PPO with clip 0.2, discount $\gamma = 0.99$, GAE $\lambda = 0.95$; control rate 100 Hz; reward weights $w_c = 0.5$, $w_e = 10^{-3}$; force-latent auxiliary weights $(\lambda_w, \lambda_s, \lambda_k, \lambda_m)$ as in Eq. (3).

Algorithm 1 COMPLIANTWBC training pipeline

Require: Motion dataset $\mathcal{M}$, force-sampling distribution $\mathcal{C}$ (§4.2), simulator $\mathcal{S}$

Stage 1: Force-aware base policy (PPO) + Force Latent Encoder (aux).

Initialize π_{base}, encoder ϕ, wrench head ψ, projection head g randomly.
for $i = 1, \dots, N_1$ **do**
  Sample $(m, \boldsymbol{f}_{\text{ext}}, y, \mathcal{C}_s) \sim \mathcal{M} \times \mathcal{C}$ (y: perturbation class).
  for $t = 1, \dots, T$ **do**
    $(\boldsymbol{\mu}_t, \log \boldsymbol{\sigma}_t^2) \leftarrow \phi(\boldsymbol{q}_{t-H:t}, \boldsymbol{\nu}_{t-H:t}, \boldsymbol{\tau}_{t-H:t-1}, \boldsymbol{c}_{t-H:t})$; $\boldsymbol{z}_F \leftarrow \boldsymbol{\mu}_t$.detach().
    $\boldsymbol{\tau}_t \leftarrow \pi_{\text{base}}(\boldsymbol{o}_t, \boldsymbol{c}_t, \boldsymbol{z}_F)$; step $\mathcal{S}$.
    Roll the analytical controller on the same state to obtain $\boldsymbol{x}^{\text{c}}_{\ell,d}$; compute reward (2).
  end for
  PPO update on π_{base} [44].
  Aux update on (ϕ, ψ, g): $\nabla \mathcal{L}_\phi$ from Eq. (3) using $(\boldsymbol{f}_{\text{ext}}, y)$.
end for

Stage 2: Residual on impedance targets (PPO).

Freeze π_{base}, ϕ, ψ. Initialize the residual policy π_{res} randomly.
for $i = 1, \dots, N_2$ **do**
  Sample $(m, \boldsymbol{f}_{\text{ext}}, \mathcal{C}_s) \sim \mathcal{M} \times \mathcal{C}$.
  for $t = 1, \dots, T$ **do**
    $\boldsymbol{z}_F \leftarrow \phi(\cdot)$; $\hat{\boldsymbol{f}}_{\text{ext}} \leftarrow \psi(\boldsymbol{z}_F)$.
    Form analytical target $\boldsymbol{x}_{\ell,d}$ via Eq. (1) using $\hat{\boldsymbol{f}}_{\text{ext}}$.
    $\Delta \boldsymbol{x}_{\ell,d} \leftarrow \pi_{\text{res}}(\boldsymbol{o}_t, \boldsymbol{c}_t, \boldsymbol{z}_F, \hat{\boldsymbol{f}}_{\text{ext}})$.
    Compose $\tilde{\boldsymbol{x}}_{\ell,d}$ via (4); $\boldsymbol{K}_\ell$ held fixed by curriculum.
    $\boldsymbol{\tau}_t \leftarrow \pi_{\text{base}}(\boldsymbol{o}_t, \boldsymbol{c}_t, \boldsymbol{z}_F)$ executed against $(\tilde{\boldsymbol{x}}_{\ell,d}, \boldsymbol{K}_\ell)$; step $\mathcal{S}$; compute reward (2).
  end for
  PPO update on π_{res} [44].
end for
return $\pi_{\text{base}}, \phi, \psi, \pi_{\text{res}}$.

E Metrics Formulation

Task success. A trial is successful if the robot remains valid throughout the force perturbation rollout without triggering the evaluation termination conditions. In our evaluation, these terminations are restricted to fall/instability events: root linear velocity exceeding $5\,\text{m/s}$, root height dropping below $0.5\,\text{m}$, or root height rising above $1.0\,\text{m}$. Let $f_i \in \{0, 1\}$ denote whether any such termination is triggered during trial i. The success indicator is

$$s_i = \mathbf{1}[f_i = 0], \tag{16}$$

and the reported success rate is

$$S = \frac{1}{N} \sum_{i=1}^{N} s_i. \tag{17}$$

Command tracking error. For the commanded body frames $\mathcal{K}_{\text{cmd}}$, currently the left and right wrist frames, we measure Cartesian tracking error relative to the reference motion. Let $\mathcal{T}_{\text{free}}$ denote timesteps for which the external-force envelope is inactive, and let $\mathcal{T}_{\text{force}}$ denote timesteps for which it is active, i.e., the normalized applied-force envelope $a_t > 0.05$ (the trapezoid amplitude of §4.2). For a time set $\mathcal{T}$, the command tracking RMSE is

$$E_{\text{cmd}}(\mathcal{T}) = \sqrt{\frac{1}{|\mathcal{T}|\,|\mathcal{K}_{\text{cmd}}|} \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}_{\text{cmd}}} \left\| \boldsymbol{p}_{k,t} - \boldsymbol{p}^{\text{ref}}_{k,t} \right\|_2^2}. \tag{18}$$

We report the free-space error $E^{\text{free}}_{\text{cmd}} = E_{\text{cmd}}(\mathcal{T}_{\text{free}})$ from the no-force pass and the force-window error $E^{\text{force}}_{\text{cmd}} = E_{\text{cmd}}(\mathcal{T}_{\text{force}})$ over the active perturbation window, which quantifies tracking relaxation under external contact.

Table 3: Full reward bundle for COMPLIANTWBC (both stages), grouped by purpose with per-term weights; the compliance-fidelity term of Eq. (2) corresponds to the *Compliance* group.

Term	Form	Weight
Whole-body tracking		
`tracking_joint_dof`	$\exp\big(-0.15 \sum_j (q_j - q_{j,d})^2\big)$	+2.0
`tracking_joint_vel`	$\exp\big(-0.01 \sum_j (\dot{q}_j - \dot{q}_{j,d})^2\big)$	+0.2
`tracking_root_rotation`	$\exp(-5\, \theta_q^2)$	+1.0
`tracking_root_linear_vel`	$\exp(-\lVert \mathbf{v}_b - \mathbf{v}_{b,d} \rVert^2)$	+1.0
`tracking_root_angular_vel`	$\exp(-\lVert \boldsymbol{\omega}_b - \boldsymbol{\omega}_{b,d} \rVert^2)$	+1.0
`tracking_keybody_pos`	$\exp\big(-10 \sum_k \lVert \Delta_b^k - \Delta_{b,d}^k \rVert^2\big)$	+2.0
`tracking_keybody_pos_global`	$\exp\big(-10 \sum_k \lVert \mathbf{x}_k - \mathbf{x}_{k,d} \rVert^2\big)$	+0.1
Compliance — admittance force & virtual target		
`force_reward`	$\exp\big(-0.0625\, \langle \lVert \mathbf{F}_\ell^{\text{app}} - \mathbf{F}_\ell^{\text{exp}} \rVert \rangle_\ell^2\big)$ ⊮[not exc.]	+2.0
`force_target_tracking`	$\exp\big(-25\, \langle \lVert \mathbf{x}_\ell - \tilde{\mathbf{x}}_\ell \rVert \rangle_\ell^2\big)$	+2.0
`force_target_vel_tracking`	$\exp\big(-\langle \lVert \dot{\mathbf{x}}_\ell - \dot{\tilde{\mathbf{x}}}_\ell \rVert \rangle_\ell^2\big)$	+1.0
`force_exd_penalty`	$\langle$⊮$[\lVert \mathbf{F}_\ell^{\text{app}} \rVert > F_{\max,\ell} \wedge \lVert \mathbf{F}_\ell^{\text{app}} \rVert > \lVert \mathbf{F}_\ell^{\text{exp}} \rVert + \frac{\delta}{2}]\rangle_\ell$	−6.0
`keypoint_tracking_imp`	$\exp\big(-10 \sum_{k \in \mathcal{K}} \lVert \mathbf{x}_k - \tilde{\mathbf{x}}_k \rVert^2\big),\ \tilde{\mathbf{x}}_k = \mathbf{x}_\ell^{\text{virt}}$ on $\mathcal{L}$	+2.0
Lower-body yielding under load		
`lower_keypoint_tracking`	$\exp\big(-10 \sum_{k \in \mathcal{K}_{\text{lo}}} \lVert \mathbf{x}_k - \tilde{\mathbf{x}}_{k,\text{lo}} \rVert^2\big)$	+2.0
`tracking_compliant_root_translation_z`	$\exp\big(-5\, (z_{\text{root}} - z^*_{\text{virt}})^2\big)$	+1.0
`squat_reward`	$\exp\big(-25\, (\Delta z - c_z \sum_\ell \lVert \mathbf{F}_\ell \rVert)^2\big),\ c_z = 3 \times 10^{-4}$ m/N	+1.0
Safety — joint, torque, contact, collision		
`alive`	⊮[alive]	+0.5
`dof_pos_limits`	$\sum_j \big[\text{ReLU}(q_j - q_j^{\max}) + \text{ReLU}(q_j^{\min} - q_j)\big]$	−5.0
`joint_limit`	same form as `dof_pos_limits` (additional)	−10.0
`dof_torque_limits`	$\sum_j \text{ReLU}\big(\lvert \tau_j \rvert / \tau_{j,\max} - 0.95\big)$	−1.0
`self_collisions`	⊮$[\, \lVert \mathbf{F}^{\text{self}} \rVert > 10\text{ N}\,]$ summed over contact pts	−10.0
`feet_stumble`	⊮$[\exists f : \lVert \mathbf{F}_f^{xy} \rVert > 4 \lvert F_f^z \rvert]$	−1.25
`feet_contact_forces`	$\sum_f \text{ReLU}\big(\lvert F_f^z \rvert - 500\text{ N}\big)$	-5×10^{-4}
`feet_slip`	$\sum_f \sqrt{\lVert \dot{\mathbf{x}}_f^{xy} \rVert}$ ⊮$[\lvert F_f^z \rvert > 5]$	−0.1
Regularization — smoothness & locomotion		
`dof_vel`	$\sum_j \dot{q}_j^2$	-10^{-4}
`dof_acc`	$\sum_j \ddot{q}_j^2$	-5×10^{-8}
`ankle_dof_vel`	$\sum_{j \in \mathcal{J}_{\text{ank}}} \dot{q}_j^2$	-2×10^{-4}
`ankle_dof_acc`	$\sum_{j \in \mathcal{J}_{\text{ank}}} \ddot{q}_j^2$	-10^{-7}
`action_rate_l2`	$\lVert \mathbf{a}_t - \mathbf{a}_{t-1} \rVert^2$	−0.1
`ang_vel_xy`	$\omega_{b,x}^2 + \omega_{b,y}^2$	−0.01
`feet_air_time`	$\sum_f \min(0, t_f^{\text{air}} - 0.5\text{s})$ ⊮[first contact] ⊮$[\lVert \mathbf{v}_{b,d}^{xy} \rVert > 0.05]$	+5.0

Torque reaction. We report the fraction of near-saturated joints, ρ_τ:

$$\rho_\tau = \tfrac{1}{T n_j} \sum_{t=1}^{T} \sum_{j=1}^{n_j} \mathbf{1}\left[\lvert \tau_{j,t} \rvert > 0.9 \tau_j^{\max}\right], \tag{19}$$

where $\tau_j^{\max}$ is the torque limit of joint j. A stiff controller fights the force (high ρ_τ); a compliant one yields, keeping it low.

Lower-body participation. To verify that the controller actually uses the lower body under multi-site contacts, we compare the torque deviation from a matched no-perturbation rollout:

$$R_{\text{LB}} = \frac{\sum_t \sum_{j \in \mathcal{J}_{\text{lower}}} \lvert \tau_{j,t} - \tau_{j,t}^0 \rvert}{\sum_t \sum_{j \in \mathcal{J}_{\text{all}}} \lvert \tau_{j,t} - \tau_{j,t}^0 \rvert + \epsilon}. \tag{20}$$

Here $\boldsymbol{\tau}^0$ is the torque sequence produced by the same controller on the same command trajectory without external perturbation. A high R_{LB} during hip, pelvis, torso, or knee perturbations indicates that the policy redistributes effort through the legs instead of treating force response as an upper-body-only problem.

F Local passivity by construction

Proposition 1 (Bounded interaction force and local passivity)**.** *Let the closed-loop system satisfy: (i) the base policy realizes* $\boldsymbol{\tau}_{\text{c}}$ *of Eq. (7) at the equilibrium* $(\boldsymbol{q}^\star, \mathbf{0})$*; (ii) the residual output is bounded* *as in Eq. (4); (iii) the per-link stiffness $\boldsymbol{K}_\ell$ is constant in time. Then (a) the residual shifts each link equilibrium by at most ϵ_x, so the induced restoring force is bounded by $\|\boldsymbol{K}_\ell\|\,\epsilon_x$; and (b) if additionally (iv) the residual-edited equilibrium $\tilde{\boldsymbol{x}}_{\ell,d}$ varies slowly enough that its power injection is dominated by the link damping $\boldsymbol{D}_\ell$, the storage function $S = \frac{1}{2}\boldsymbol{\nu}^\top \boldsymbol{M}\boldsymbol{\nu} + \frac{1}{2}\sum_\ell (\boldsymbol{x}_\ell - \tilde{\boldsymbol{x}}_{\ell,d})^\top \boldsymbol{K}_\ell (\boldsymbol{x}_\ell - \tilde{\boldsymbol{x}}_{\ell,d})$ satisfies $\dot{S} \leq \sum_i \boldsymbol{f}_{ext,i}^\top \dot{\boldsymbol{x}}_{ext,i}$, where $\boldsymbol{x}_{ext,i}$ is the displacement of contact point i; i.e. the closed loop is passive with respect to the external-wrench port.*

Sketch. Bound (a) is immediate from Eq. (4): the residual moves the equilibrium by at most ϵ_x, so with fixed $\boldsymbol{K}_\ell$ the induced restoring force changes by at most $\|\boldsymbol{K}_\ell\|\,\epsilon_x$. For (b), differentiating S and substituting the closed-loop dynamics gives $\dot{S} = \sum_i \boldsymbol{f}_{\text{ext},i}^\top \dot{\boldsymbol{x}}_{\text{ext},i} - (\text{damping dissipation}) - \sum_\ell (\boldsymbol{x}_\ell - \tilde{\boldsymbol{x}}_{\ell,d})^\top \boldsymbol{K}_\ell \,\dot{\tilde{\boldsymbol{x}}}_{\ell,d}$. With $\boldsymbol{K}_\ell$ constant the $\dot{\boldsymbol{K}}_\ell$ term that would otherwise inject non-passive power vanishes identically; the remaining moving-equilibrium term is dominated by the damping dissipation under assumption (iv), yielding the stated inequality. For the general time-varying-stiffness case the energy tank of §A would absorb the $\dot{\boldsymbol{K}}_\ell$ term; in our pipeline stiffness is fixed and the residual does not modulate gains, so the tank is reference scaffolding rather than a runtime mechanism.

Scope and what this does and does not buy. The result is *local* (around the equilibrium) and *by construction*: it follows from the residual bound and fixed stiffness, not from the learned components. Global stability under adversarial external forces is not claimed and remains open. Two consequences are worth stating explicitly. (i) Action-residual baselines (ASAP, ResMimic) edit the torque directly and do not admit this storage function; their effect on induced contact force is not physically bounded by construction. (ii) The proposition is a safety property for human–robot contact, not a performance property: it does not guarantee tracking, success rate, or sim-to-real transfer. Those are settled empirically (§5).