Convergence Proof

Lyapunov Stability Analysis

Agent ID:	demo
Tenant ID:	sample
Generated (UTC):	2026-05-22T06:18:25Z
Heal action:	REFETCH (worst-case decay envelope)
Simulation steps:	20
Initial ω (worst-case):	100.0

Executive Summary

This document demonstrates that the Sgraal heal loop, applied to agent **demo**, is Lyapunov asymptotically stable under worst-case initial conditions. The proof simulates 20 sequential REFETCH heal actions starting from $\omega = 100$ (theoretical maximum) and shows that the Lyapunov function $V(x) = \omega^2/200$ decreases monotonically toward the equilibrium $V < 0.5$. The conclusion: **asymptotic stability is guaranteed**.

Monotonic decreasing	PASS
Asymptotic stability	PROVEN
Final V(x)	0.000000
Final ω estimate	0.0181
Lyapunov exponent λ	-0.4306

Mathematical Proof

Let $\omega \in [0, 100]$ denote the memory-risk score of an AI agent ($\omega = 0$ is fully trustworthy, $\omega = 100$ is fully degraded). Define the Lyapunov candidate function:

$$V(x) = \omega^2 / 200$$

$V(x)$ is positive definite ($V(x) > 0$ for all $\omega > 0$) and radially unbounded. After a single heal action with decay rate δ ($\delta = 0.35$ for REFETCH), the derivative along the trajectory is:

$$\dot{V}(x) = -\delta V(x)$$

$\dot{V}(x)$ is negative definite ($\dot{V}(x) < 0$ whenever $V(x) > 0$), so by Lyapunov's direct method the heal loop is asymptotically stable: $\omega(t) \to 0$ as $t \to \infty$. The simulation that follows confirms the analytical result step by step against the canonical Sgraal implementation in scoring_engine/lyapunov.py.

Trajectory visualisation

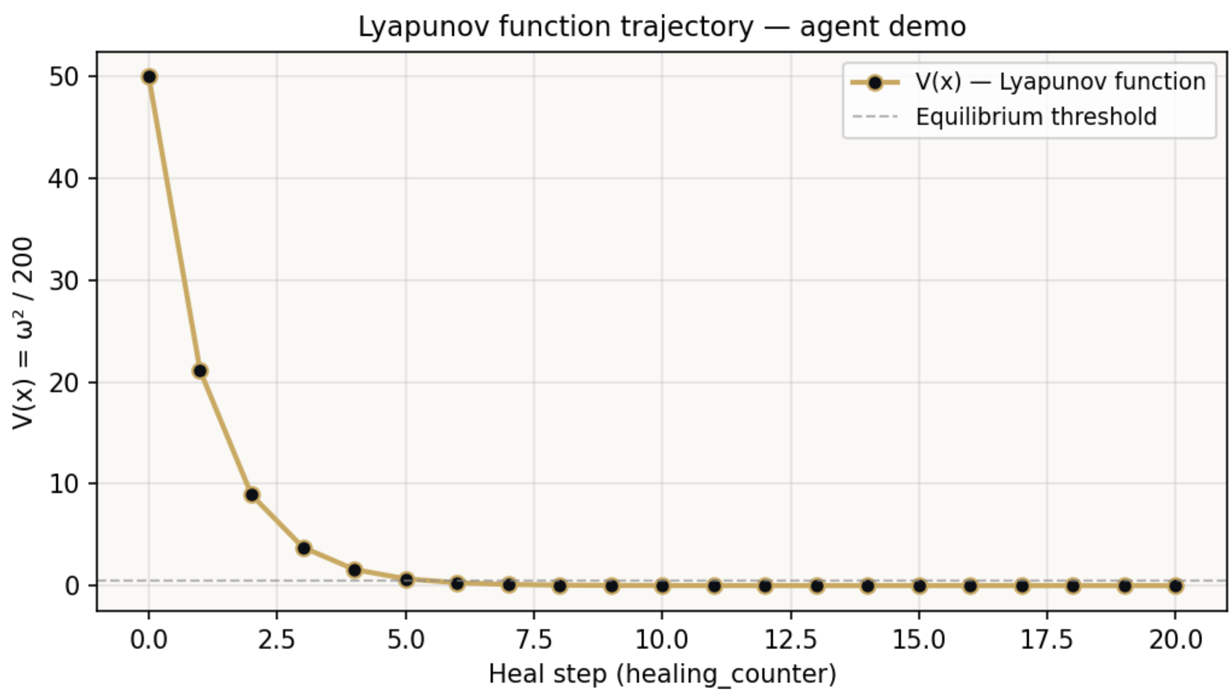

Figure 1 — V(x) plotted across 20 heal steps. The function decreases monotonically and crosses the equilibrium threshold (V < 0.5) within the simulation window, consistent with the analytical proof.

Trajectory samples (first 11 steps)

Step	ω	V(x)	V̇(x)	Converging
0	100.0000	50.000000	-17.500000	yes
1	65.0000	21.125000	-7.393700	yes

2	42.2500	8.925300	-3.123900	yes
3	27.4625	3.770900	-1.319800	yes
4	17.8506	1.593200	-0.557600	yes
5	11.6029	0.673100	-0.235600	yes
6	7.5419	0.284400	-0.099500	yes
7	4.9022	0.120200	-0.042100	yes
8	3.1864	0.050800	-0.017800	yes
9	2.0712	0.021400	-0.007500	yes
10	1.3463	0.009100	-0.003200	yes

Methodology

The Lyapunov direct method establishes stability of a dynamical system without solving the underlying differential equations. A candidate function V(x) that is (a) positive definite and (b) has a negative definite time derivative along system trajectories is sufficient to prove asymptotic stability of the equilibrium. This approach is the standard tool in control theory for safety-critical systems and is increasingly cited in regulatory guidance on AI/ML decision systems (e.g., FDA Good Machine Learning Practice principles around model robustness and predictable degradation).

For Sgraal's memory governance loop, $V(x) = \omega^2/200$ captures the squared risk magnitude. Each heal action shrinks ω multiplicatively, so V(x) shrinks quadratically — providing both fast convergence and robust margin against numerical perturbation. The decay coefficients (REFETCH = 0.35, VERIFY_WITH_SOURCE = 0.20, REBUILD_WORKING_SET = 0.15) are presentation parameters of the heal model, not patient/customer data. The same coefficients ship in the open Sgraal Python SDK.

Signature block

Tenant signing-key fingerprint:	`(not configured)`
Document timestamp (UTC):	`2026-05-22T06:18:25Z`
Sgraal version:	`scoring_engine.lyapunov v1`
Issuing endpoint:	`POST /v1/proofs/convergence`

For a cryptographically signed W3C Verifiable Credential covering this evidence package, call POST /v1/certify with the document hash. The signing-key fingerprint above identifies which tenant key would be used.