Demo Factor Analysis
(lbfgs.csv)

Sage

September 23, 2025

Data Matrix

Columns used: `time(ns)`, `memory(MB)`, `nvmops`, `neval_obj`, `neval_grad`.
Settings: `mem` and `nvar` are not inputs to PCA/FA; so I ignored them for a while.

Using $k = 2$ as number of factors

The PCA results are:

Table 1: Explained variance ratio (EVR) and cumulative EVR.

PC	EVR	Cumulative
1	0.591	0.591
2	0.280	0.871
3	0.117	0.989
4	0.010	0.999
5	0.001	1.000

Since $R_2 = 0.87126599 > 0.80$, the first two PCs already cover the vast majority of variance. Hence we retain $k = 2$ for downstream FA to balance fidelity and interpretability.

PCA Loadings (PC1 & PC2)

Table 2: PCA loadings for the first two components.

Variable	PC1	PC2
time(ns)	0.576	0.058
memory(MB)	0.571	0.008
nvmops	0.575	0.096
neval_obj	−0.104	0.691
neval_grad	−0.031	0.714

Method (PCA → FA)

- Standardize metrics (mean 0, var 1) to avoid scale dominance.
- PCA: decide dimensionality k via R_k; keep $k = 2$.
- Factor Analysis (FA) with $k = 2$ to obtain interpretable latent factors and per-run factor scores.

Results

Loadings (FA, $k = 2$). Factor 1 loads on time/memory/nvmops (resource/efficiency), Factor 2 loads on eval counts (calls).

Table 3: Factor loadings (variables × factors).

Variable	Factor 1	Factor 2
time(ns)	0.992	−0.043
memory(MB)	0.959	−0.125
nvmops	1.000	0.003
neval_obj	−0.083	0.497
neval_grad	0.038	0.849

Factor scores (first 10 rows). Lower is better on both dimensions.

Table 4: Per-run factor scores (subset).

Row	F1	F2	$S = F1 + F2$
0	−0.232	−0.451	−0.684
1	0.079	2.501	2.580
2	−0.017	2.342	2.325
3	−0.001	2.346	2.345
4	−0.025	2.245	2.220
5	−0.005	2.211	2.205
6	−0.052	2.170	2.118
7	−0.025	2.150	2.125
8	−0.056	2.100	2.044
9	−0.063	2.054	1.992

Demo Score Function

We use a simple aggregate:

$$S \;=\; F_1 \;+\; F_2 \quad \text{(lower is better)}.$$

Rationale: equal-weight combination of resource/efficiency and evaluation-call factors; easy to interpret and revise later.