Xopt NSGA-II Benchmarking

Christopher M. Pierce

March 3, 2025

Summary

This note contains benchmarking results for the new Xopt NSGA-II implementation comparing it to Xopt's "CNSGA" generator and an alternative implementation of NSGA-II from the library pymoo. The optimization algorithms were run on benchmark problems from the WFG suite with varying number of decision variables. Each of the algorithms was set to a population size of 50, crossover probability of 50%, mutation rate of $1/n$ where n is the number of decision variables and $\eta_m = \eta_c = 20$.

Note that in Xopt CNSGA, which uses operators from DEAP, the mutation and crossover probabilities are not the typical definitions. The per variable mutation probability as well as variation probability and exchange probability (or δ_1 and δ_2) are all hardcoded in the DEAP operators. These values become equal to the ones used in the other algorithms here when the CNSGA parameters "mutation_probability" and "crossover_probability" are equal to one.

Every algorithm was run 32 times on each problem. The tables below show the mean value of the IGD metric for all runs after the listed number of generations has passed. The standard deviation of the values is included in parenthesis. The values are compared to the new Xopt NSGA-II implementation using the Wilcoxon rank sum test with a significance level of 0.05. A "+", "-", or "≈" symbol are shown in the cell depending on whether it performed better, worse, or equivalent to the Xopt NSGA-II implementation up to the significance level. At the bottom of the table, the numbers of each symbol for that algorithm are aggregated.

Table 1: Summary of statistical results of 32 runs of each algorithm on WFG problems after 10 generations.

Problem	n	pymoo nsgaii	xopt cnsga	xopt nsgaii
WFG1	16	1.908e+0 (8.186e-2) +	1.945e+0 (6.625e-2) ≈	1.942e+0 (7.914e-2)
	32	1.937e+0 (9.100e-2) ≈	1.998e+0 (7.092e-2) -	1.956e+0 (8.129e-2)
	64	1.960e+0 (7.420e-2) ≈	2.004e+0 (7.271e-2) -	1.957e+0 (7.528e-2)
WFG2	16	4.861e-1 (3.771e-2) -	4.913e-1 (3.819e-2) -	4.585e-1 (5.096e-2)
	32	5.550e-1 (4.231e-2) ≈	5.724e-1 (3.984e-2) -	5.722e-1 (9.592e-2)
	64	6.236e-1 (2.841e-2) ≈	6.213e-1 (3.227e-2) ≈	6.109e-1 (3.513e-2)
WFG3	16	3.899e-1 (3.289e-2) -	3.792e-1 (2.290e-2) -	3.538e-1 (3.475e-2)
	32	4.968e-1 (2.991e-2) -	5.023e-1 (2.606e-2) -	4.774e-1 (2.637e-2)
	64	5.830e-1 (1.385e-2) -	5.839e-1 (1.811e-2) -	5.723e-1 (1.710e-2)
WFG4	16	2.509e-1 (3.788e-2) -	2.428e-1 (3.808e-2) ≈	2.381e-1 (3.886e-2)
	32	3.153e-1 (3.380e-2) ≈	3.288e-1 (3.063e-2) -	3.056e-1 (3.235e-2)
	64	3.696e-1 (2.776e-2) -	3.706e-1 (2.952e-2) -	3.515e-1 (2.698e-2)
WFG5	16	3.594e-1 (2.498e-2) -	3.663e-1 (3.124e-2) -	3.351e-1 (4.044e-2)
	32	4.599e-1 (2.406e-2) -	4.695e-1 (2.180e-2) -	4.414e-1 (3.085e-2)
	64	5.345e-1 (1.694e-2) ≈	5.407e-1 (2.554e-2) -	5.309e-1 (2.706e-2)
WFG6	16	4.366e-1 (3.808e-2) ≈	4.502e-1 (3.325e-2) -	4.231e-1 (3.477e-2)
	32	5.712e-1 (2.441e-2) -	5.762e-1 (2.627e-2) -	5.396e-1 (4.064e-2)
	64	6.650e-1 (2.065e-2) -	6.657e-1 (2.080e-2) -	6.385e-1 (2.483e-2)
WFG7	16	3.266e-1 (3.071e-2) -	3.215e-1 (2.845e-2) -	2.965e-1 (3.174e-2)
	32	4.107e-1 (2.698e-2) -	4.022e-1 (2.525e-2) -	3.847e-1 (2.754e-2)
	64	4.833e-1 (2.159e-2) -	4.828e-1 (2.698e-2) -	4.630e-1 (2.362e-2)
WFG8	16	4.624e-1 (3.849e-2) -	4.774e-1 (4.632e-2) -	4.454e-1 (2.849e-2)
	32	5.031e-1 (3.390e-2) -	4.948e-1 (2.440e-2) -	4.790e-1 (2.676e-2)
	64	5.377e-1 (1.945e-2) -	5.377e-1 (2.591e-2) -	5.196e-1 (3.087e-2)
WFG9	16	3.548e-1 (4.823e-2) -	3.643e-1 (5.892e-2) -	3.078e-1 (4.630e-2)
	32	5.120e-1 (5.938e-2) -	5.151e-1 (4.478e-2) -	4.659e-1 (7.683e-2)
	64	6.593e-1 (4.950e-2) -	6.359e-1 (4.589e-2) -	6.108e-1 (4.623e-2)
+/-/≈		1/19/7	0/24/3	

Table 2: Summary of statistical results of 32 runs of each algorithm on WFG problems after 30 generations.

Problem	**n**	**pymoo nsgaii**	**xopt cnsga**	**xopt nsgaii**
WFG1	16	1.684e+0 (5.395e-2) ≈	1.745e+0 (5.158e-2) -	1.706e+0 (5.570e-2)
	32	1.737e+0 (4.064e-2) ≈	1.798e+0 (5.685e-2) -	1.750e+0 (6.028e-2)
	64	1.777e+0 (5.617e-2) ≈	1.850e+0 (6.379e-2) -	1.777e+0 (7.252e-2)
WFG2	16	3.054e-1 (5.417e-2) ≈	2.869e-1 (6.474e-2) ≈	2.887e-1 (6.559e-2)
	32	3.618e-1 (5.661e-2) ≈	3.646e-1 (5.706e-2) ≈	3.865e-1 (1.119e-1)
	64	4.536e-1 (4.612e-2) ≈	4.474e-1 (5.416e-2) ≈	4.410e-1 (4.810e-2)
WFG3	16	1.627e-1 (2.688e-2) ≈	1.619e-1 (2.572e-2) ≈	1.569e-1 (2.680e-2)
	32	2.716e-1 (3.059e-2) ≈	2.804e-1 (2.642e-2) -	2.645e-1 (3.164e-2)
	64	3.928e-1 (2.267e-2) ≈	4.007e-1 (2.012e-2) ≈	3.945e-1 (2.409e-2)
WFG4	16	9.378e-2 (1.140e-2) ≈	1.023e-1 (1.484e-2) -	9.617e-2 (1.115e-2)
	32	1.476e-1 (1.195e-2) ≈	1.601e-1 (1.537e-2) -	1.451e-1 (1.338e-2)
	64	2.073e-1 (1.848e-2) ≈	2.178e-1 (2.522e-2) -	2.002e-1 (1.351e-2)
WFG5	16	1.528e-1 (1.530e-2) ≈	1.619e-1 (1.752e-2) -	1.493e-1 (1.366e-2)
	32	2.342e-1 (1.579e-2) ≈	2.397e-1 (1.346e-2) -	2.293e-1 (1.769e-2)
	64	3.255e-1 (1.412e-2) ≈	3.361e-1 (1.777e-2) -	3.244e-1 (1.784e-2)
WFG6	16	1.922e-1 (3.040e-2) ≈	2.096e-1 (2.576e-2) -	1.918e-1 (2.718e-2)
	32	2.799e-1 (2.616e-2) -	2.883e-1 (2.698e-2) -	2.653e-1 (2.118e-2)
	64	3.995e-1 (1.968e-2) -	4.037e-1 (2.200e-2) -	3.799e-1 (2.560e-2)
WFG7	16	1.132e-1 (2.646e-2) ≈	1.172e-1 (2.226e-2) ≈	1.232e-1 (3.312e-2)
	32	1.974e-1 (2.876e-2) ≈	1.856e-1 (1.723e-2) +	2.020e-1 (2.754e-2)
	64	2.949e-1 (2.492e-2) ≈	2.867e-1 (1.896e-2) ≈	2.850e-1 (2.978e-2)
WFG8	16	2.748e-1 (2.305e-2) ≈	2.877e-1 (3.049e-2) -	2.750e-1 (1.923e-2)
	32	2.987e-1 (2.189e-2) ≈	2.982e-1 (1.995e-2) ≈	2.943e-1 (2.224e-2)
	64	3.473e-1 (1.584e-2) -	3.520e-1 (1.094e-2) -	3.367e-1 (2.138e-2)
WFG9	16	1.787e-1 (2.649e-2) -	1.467e-1 (3.808e-2) ≈	1.493e-1 (4.107e-2)
	32	2.139e-1 (6.132e-2) ≈	2.133e-1 (5.653e-2) ≈	2.011e-1 (4.799e-2)
	64	3.630e-1 (5.389e-2) -	3.469e-1 (3.201e-2) -	3.205e-1 (6.401e-2)
+/-/≈		0/5/22	1/16/10	

Table 3: Summary of statistical results of 32 runs of each algorithm on WFG problems after 50 generations.

Problem	**n**	**pymoo nsgaii**	**xopt cnsga**	**xopt nsgaii**
WFG1	16	1.575e+0 (7.008e-2) ≈	1.658e+0 (3.849e-2) -	1.588e+0 (8.059e-2)
	32	1.641e+0 (5.755e-2) ≈	1.709e+0 (5.611e-2) -	1.660e+0 (5.871e-2)
	64	1.684e+0 (4.339e-2) ≈	1.767e+0 (5.967e-2) -	1.694e+0 (6.181e-2)
WFG2	16	2.553e-1 (5.985e-2) ≈	2.312e-1 (7.863e-2) ≈	2.412e-1 (7.287e-2)
	32	2.946e-1 (6.489e-2) ≈	2.873e-1 (6.176e-2) ≈	3.326e-1 (1.168e-1)
	64	3.747e-1 (5.165e-2) ≈	3.569e-1 (6.161e-2) ≈	3.626e-1 (5.638e-2)
WFG3	16	1.072e-1 (2.133e-2) ≈	1.036e-1 (2.155e-2) ≈	1.045e-1 (2.345e-2)
	32	1.939e-1 (2.717e-2) ≈	1.948e-1 (2.255e-2) ≈	1.947e-1 (3.106e-2)
	64	3.018e-1 (2.557e-2) ≈	2.944e-1 (1.572e-2) ≈	2.988e-1 (2.494e-2)
WFG4	16	5.816e-2 (7.880e-3) ≈	6.296e-2 (9.746e-3) ≈	6.022e-2 (7.866e-3)
	32	9.862e-2 (7.068e-3) ≈	1.065e-1 (1.189e-2) -	9.937e-2 (8.357e-3)
	64	1.508e-1 (1.504e-2) ≈	1.570e-1 (2.323e-2) -	1.452e-1 (9.508e-3)
WFG5	16	1.059e-1 (7.039e-3) ≈	1.135e-1 (7.508e-3) -	1.051e-1 (7.510e-3)
	32	1.655e-1 (1.201e-2) ≈	1.668e-1 (8.027e-3) ≈	1.642e-1 (1.247e-2)
	64	2.409e-1 (1.411e-2) ≈	2.417e-1 (1.106e-2) ≈	2.427e-1 (1.330e-2)
WFG6	16	1.330e-1 (2.423e-2) ≈	1.444e-1 (1.878e-2) ≈	1.350e-1 (1.947e-2)
	32	1.783e-1 (1.890e-2) ≈	1.851e-1 (1.784e-2) -	1.720e-1 (1.535e-2)
	64	2.752e-1 (1.620e-2) -	2.752e-1 (1.901e-2) -	2.649e-1 (2.139e-2)
WFG7	16	5.469e-2 (1.367e-2) +	6.102e-2 (1.565e-2) ≈	7.000e-2 (2.430e-2)
	32	1.253e-1 (2.922e-2) ≈	1.086e-1 (1.088e-2) +	1.337e-1 (2.728e-2)
	64	2.085e-1 (2.748e-2) ≈	1.917e-1 (1.969e-2) +	2.099e-1 (3.313e-2)
WFG8	16	2.211e-1 (1.686e-2) ≈	2.343e-1 (1.303e-2) -	2.243e-1 (1.612e-2)
	32	2.354e-1 (1.912e-2) ≈	2.339e-1 (1.555e-2) ≈	2.351e-1 (1.915e-2)
	64	2.712e-1 (1.532e-2) ≈	2.719e-1 (1.166e-2) -	2.666e-1 (1.379e-2)
WFG9	16	1.441e-1 (4.178e-2) -	1.055e-1 (4.891e-2) ≈	1.202e-1 (5.411e-2)
	32	1.412e-1 (3.393e-2) ≈	1.409e-1 (3.607e-2) ≈	1.416e-1 (3.276e-2)
	64	2.454e-1 (4.491e-2) ≈	2.387e-1 (2.480e-2) ≈	2.345e-1 (4.960e-2)
+/-/≈		1/2/24	2/10/15	

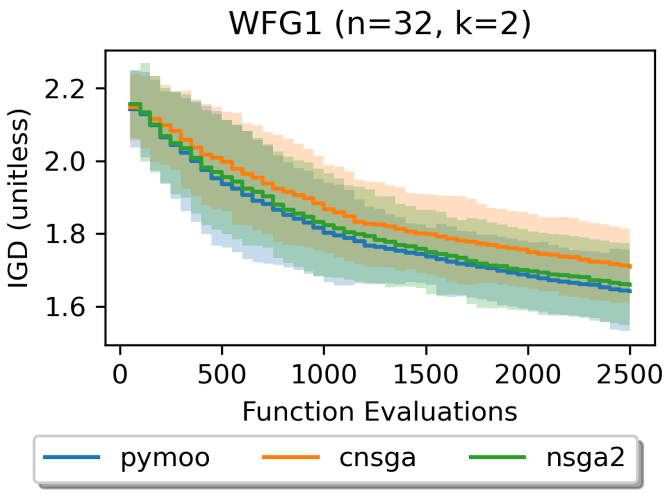

Figure 1: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG1.

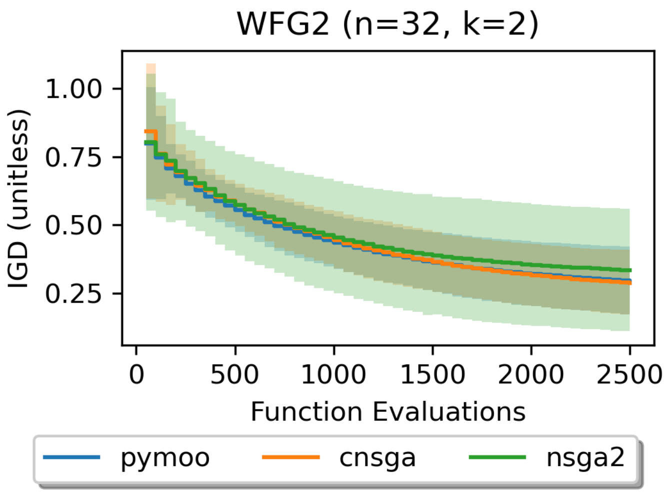

Figure 2: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG2.

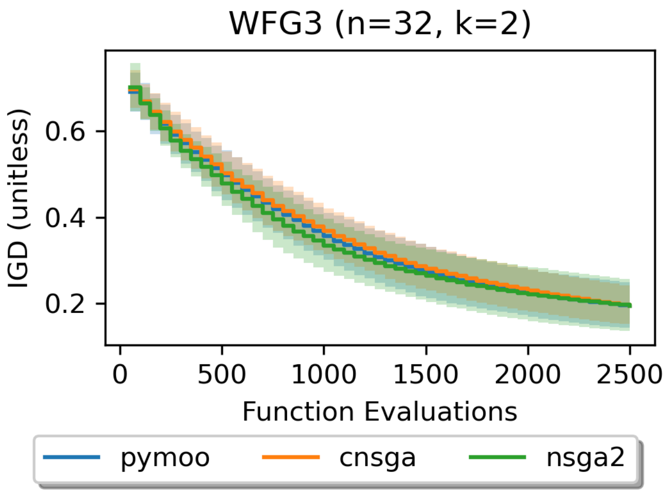

Figure 3: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG3.

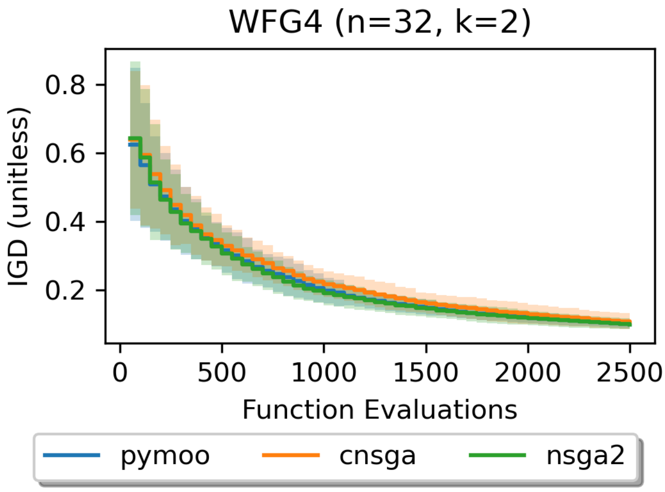

Figure 4: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG4.

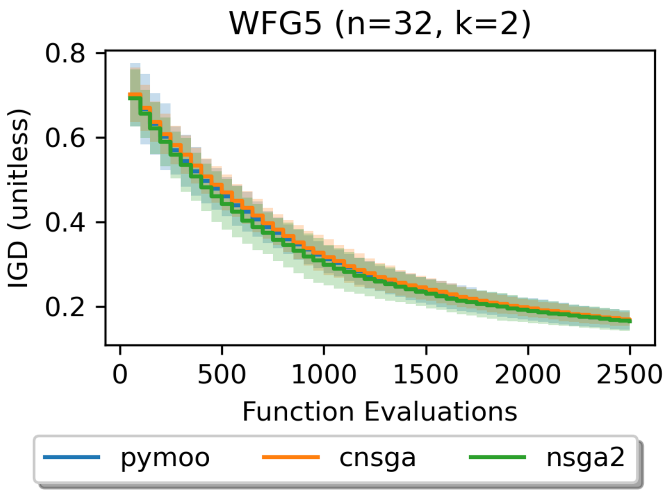

Figure 5: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG5.

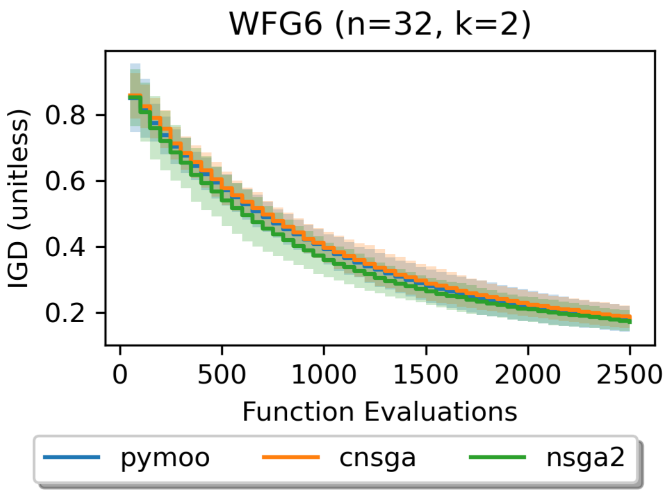

Figure 6: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG6.

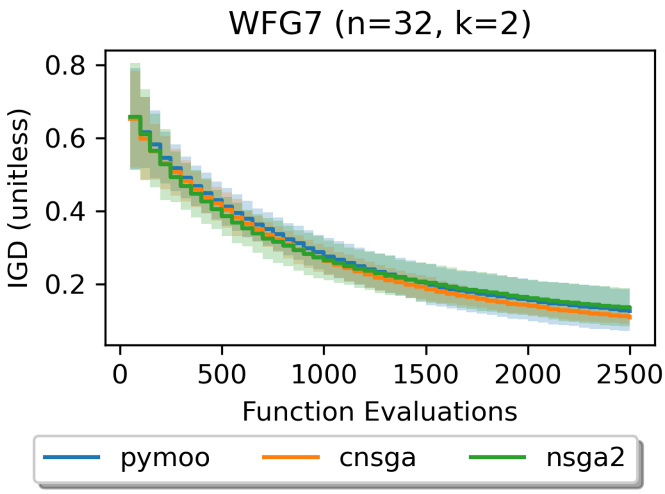

Figure 7: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG7.

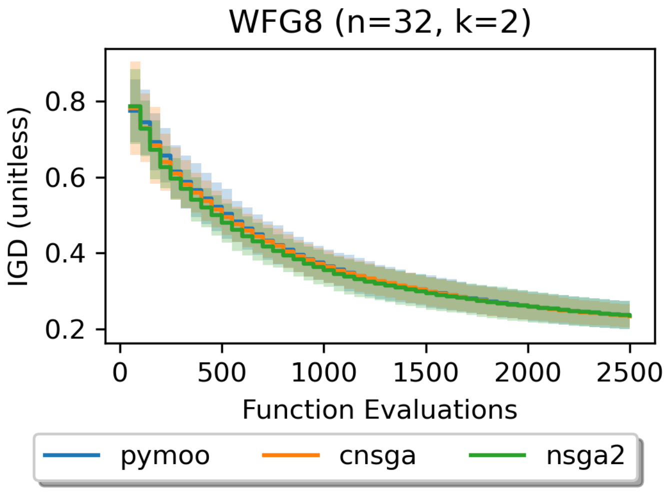

Figure 8: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG8.

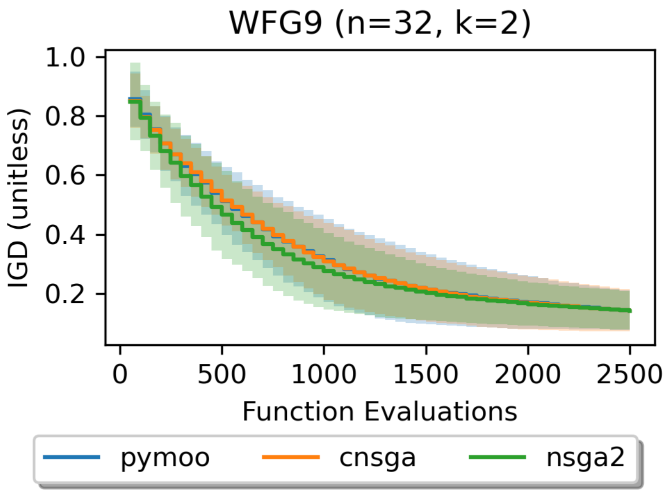

Figure 9: Mean and 95% prediction interval of the IGD metric for each algorithm on WFG9.