

AOPT Exercise 7: L-BFGS and Gauss Newton Method

SCHALLER ALAIN, FONTANA JONAS, CARREL VINCENT

Part 1

Apr

Alain, Jonas, Vincent

Exercise 7

Secant Equation:

The secant equation is

$$B_{k+1} s_k = y_k$$

where $s_k = x_{k+1} - x_k$, $y_k = \nabla f_{k+1} - \nabla f_k$, and B_{k+1} is an approximation of the Hessian $\nabla^2 f_{k+1}$.

Prove that the secant equation is valid to find the next hessian approximation B_{k+1}

We need to prove that it satisfies two conditions:

• $\nabla m_{k+1}(0) \stackrel{!}{=} \nabla f_{k+1}$:

$$\text{We have } m_k(v) = f_k + \nabla f_k^T v + \frac{1}{2} v^T B_k v$$

$$\Rightarrow \nabla m_k(v) = \nabla f_k + B_k v$$

$$\Rightarrow \nabla m_{k+1}(0) = \nabla f_{k+1} + B_{k+1} \cdot 0 = \nabla f_{k+1} \quad \checkmark$$

• $\nabla m_{k+1}(-t_k v_k) \stackrel{!}{=} \nabla f_k$:

$$\begin{aligned} \nabla m_{k+1}(-t_k v_k) &= \nabla m_{k+1}(-(x_{k+1} - x_k)) = \nabla m_{k+1}(-s_k) \\ &= \nabla f_{k+1} + B_{k+1}(-s_k) \end{aligned}$$

$$\Rightarrow B_{k+1} s_k = y_k$$

$$\Rightarrow -B_{k+1} s_k = \nabla f_k - \nabla f_{k+1}$$

$$\Rightarrow \nabla f_{k+1} + B_{k+1}(-s_k) = \nabla f_k$$

$$\Rightarrow \nabla m_{k+1}(-t_k v_k) = \nabla f_k \quad \#$$

Curvature Condition:

• Show that the curvature condition $s_k^T y_k > 0$ is required for the Quasi-Newton method

• Prove that the curvature condition is satisfied when the line search algorithm for the Wolfe conditions is used

• To ensure a descent direction, we need $B \in S^n_+$, positive definite

$$\Rightarrow B_{k+1} s_k = y_k$$

$$\Rightarrow s_k^T B_{k+1} s_k = s_k^T y_k$$

$\underbrace{s_k^T B_{k+1} s_k}_{> 0 \text{ because } B_{k+1} \text{ positive definite}} = s_k^T y_k > 0 \quad \#$

• Useful equalities:

$$v_k = -B_k^{-1} \nabla f_k$$

$$s_k = x_{k+1} - x_k = x_k - v_k B_k^{-1} \nabla f_k - x_k = -v_k B_k^{-1} \nabla f_k$$

$$\Rightarrow s_k = l_k v_k$$

$$\Rightarrow v_k = \frac{s_k}{l_k}$$

To satisfy the strong Wolfe conditions with $\beta \in (\alpha, 1)$, it must satisfy the curvature condition:

$$\nabla f(x_k + l_k v_k)^T v_k \geq \beta \nabla f(x_k)^T v_k$$

$$\Rightarrow \nabla f_{k+1}^T v_k \geq \beta \nabla f_k^T v_k$$

$$\nabla f_{k+1}^T v_k - \nabla f_k^T v_k \geq \beta \nabla f_k^T v_k - \nabla f_k^T v_k$$

$$(\nabla f_{k+1} - \nabla f_k)^T v_k \geq (\beta - 1) \nabla f_k^T v_k$$

$$y_k^T v_k \geq (\beta - 1) \nabla f_k^T v_k$$

$$y_k^T \frac{s_k}{l_k} \geq (\beta - 1) \nabla f_k^T v_k$$

$$y_k^T s_k \geq \underbrace{l_k}_{>0} \underbrace{(\beta - 1)}_{<0} \underbrace{\nabla f_k^T v_k}_{<0} > 0$$

$$\Rightarrow y_k^T s_k > 0 \quad \#$$

Programming: Method comparison

Grid size: 5

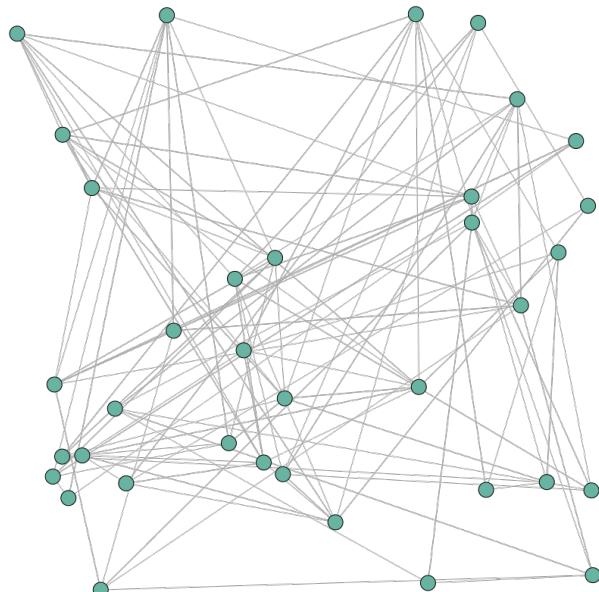


Figure 1: **Gauss Newton**: 0: f without length & Constrained spring scenario: 2: Corners – (As Gauss Newton executable uses the Newton Methods without hessian modification, and in this case, the LLT factorization fails, which makes marks the end of the method like all other *Gauss Newton without length*)

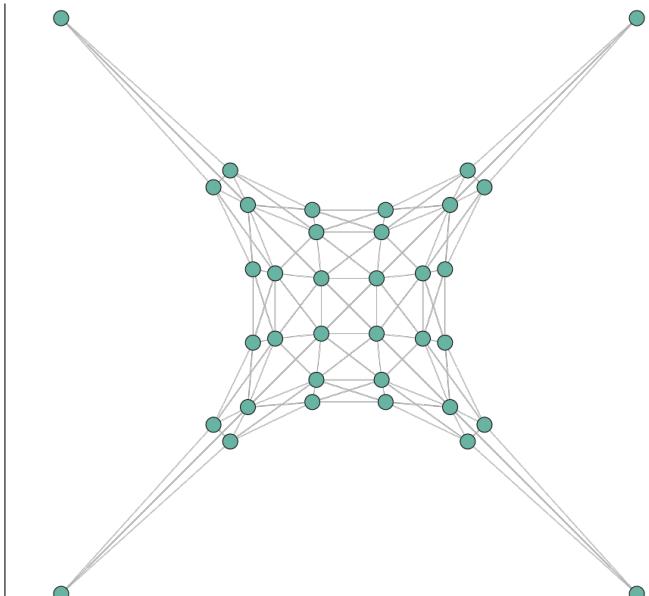


Figure 2: **Standard Newton**: 0: f without length & Constrained spring scenario: 2: Corners

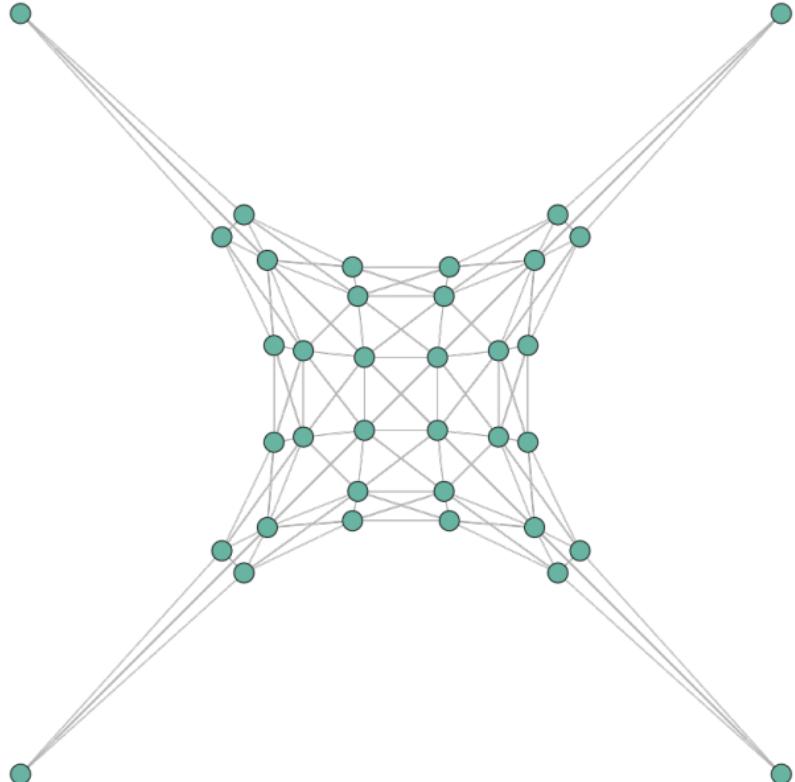


Figure 3: **Gradient Descent**: 0: f without length & Constrained spring scenario: 1: Sides

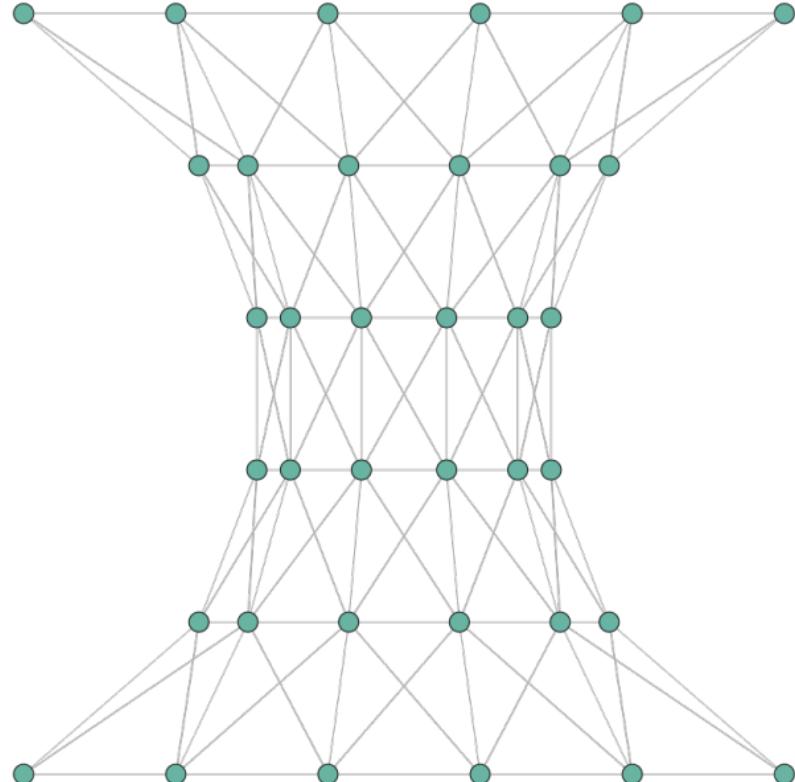


Figure 4: **Gradient Descent**: 0: f without length & Constrained spring scenario: 1: Sides

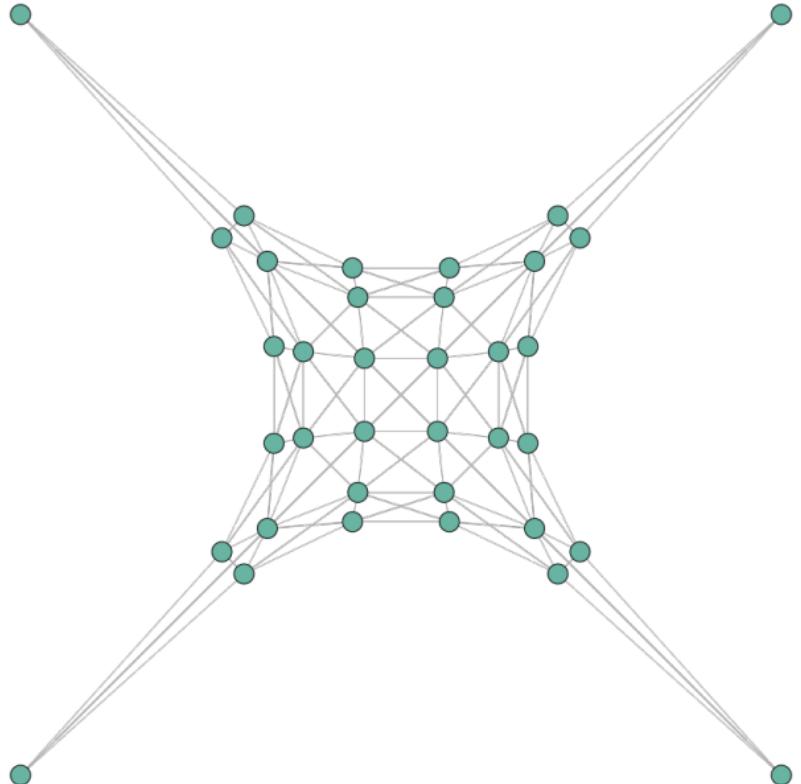


Figure 5: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

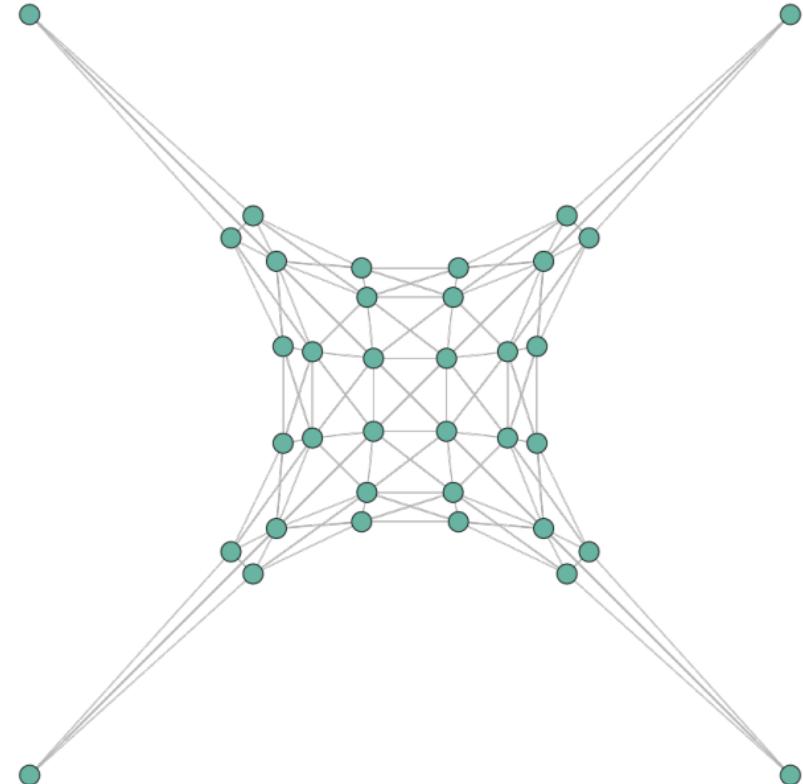


Figure 6: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

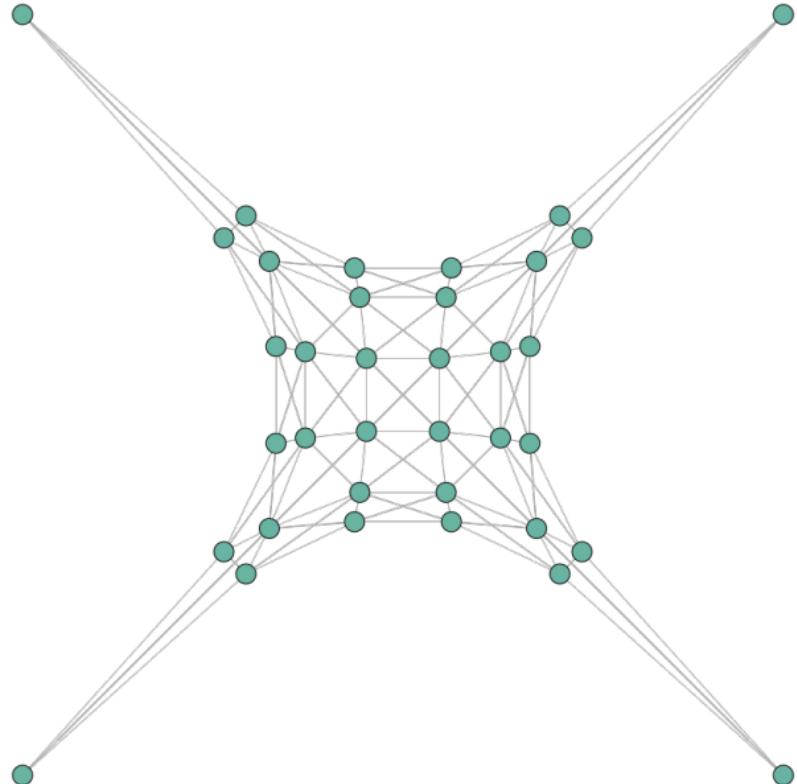


Figure 7: **L-BFGS**: 0: f without length &
Constrained spring scenario: 2: Corners

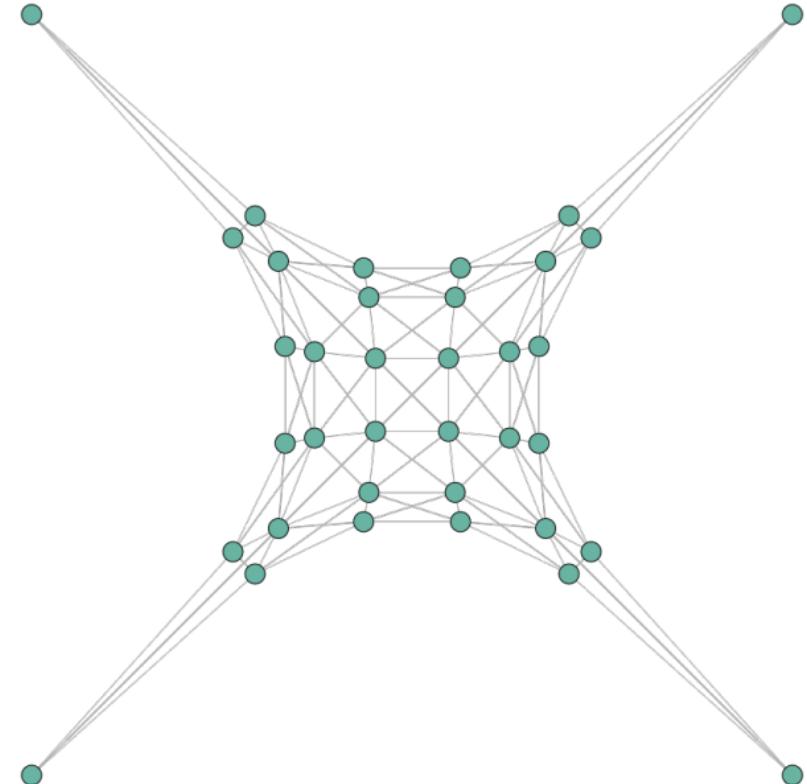


Figure 8: **L-BFGS**: 0: f without length &
Constrained spring scenario: 2: Corners

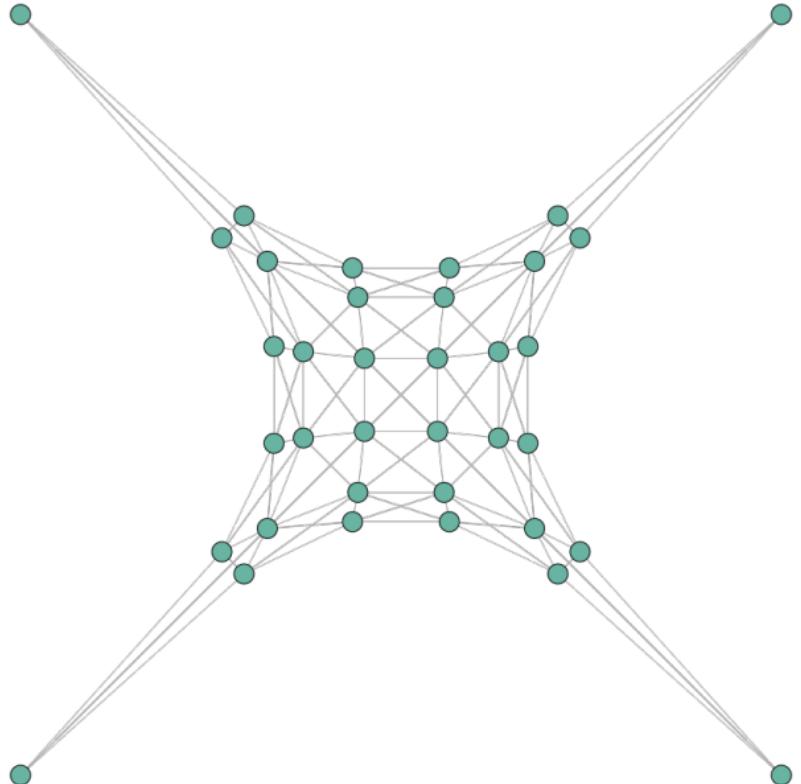


Figure 9: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

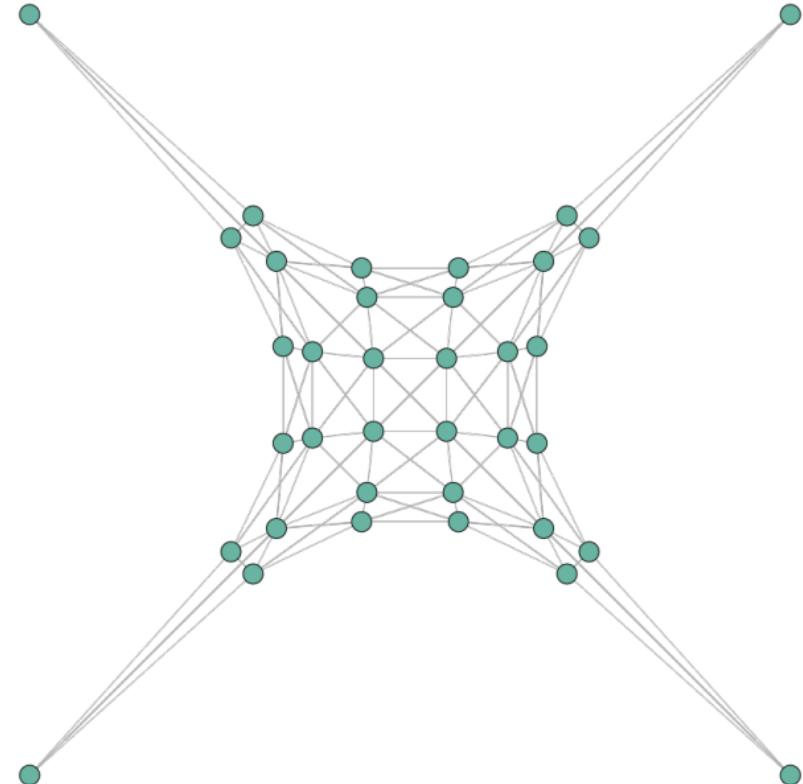


Figure 10: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

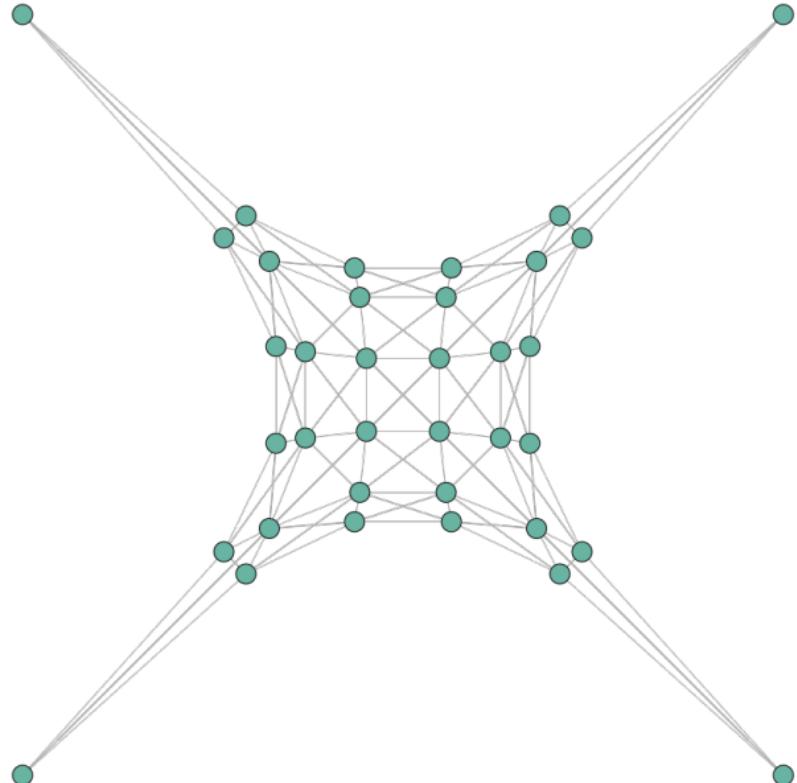


Figure 11: **L-BFGS**: 0: f without length &
Constrained spring scenario: 2: Corners

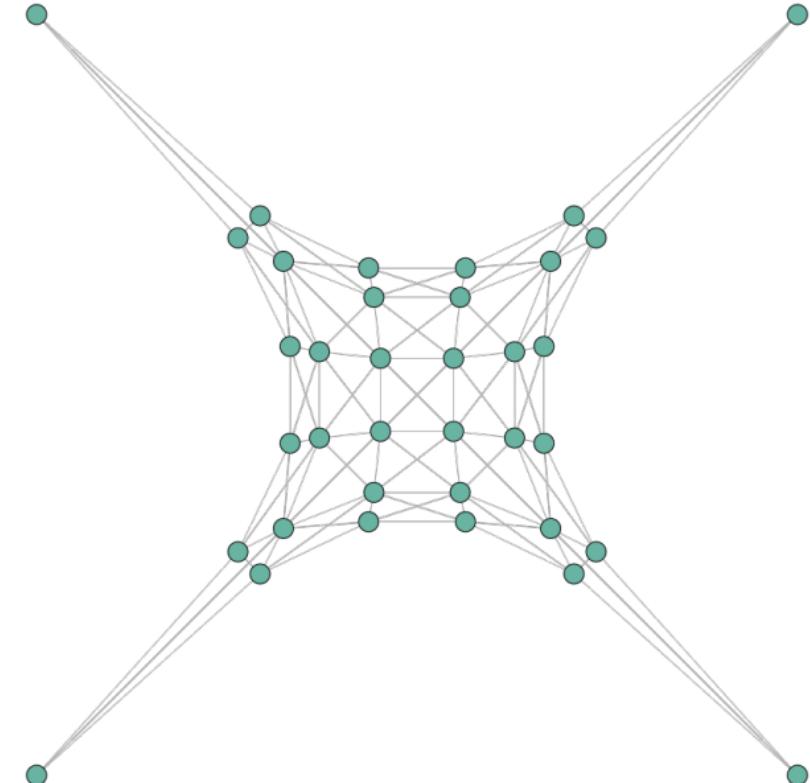


Figure 12: **L-BFGS**: 0: f without length &
Constrained spring scenario: 2: Corners

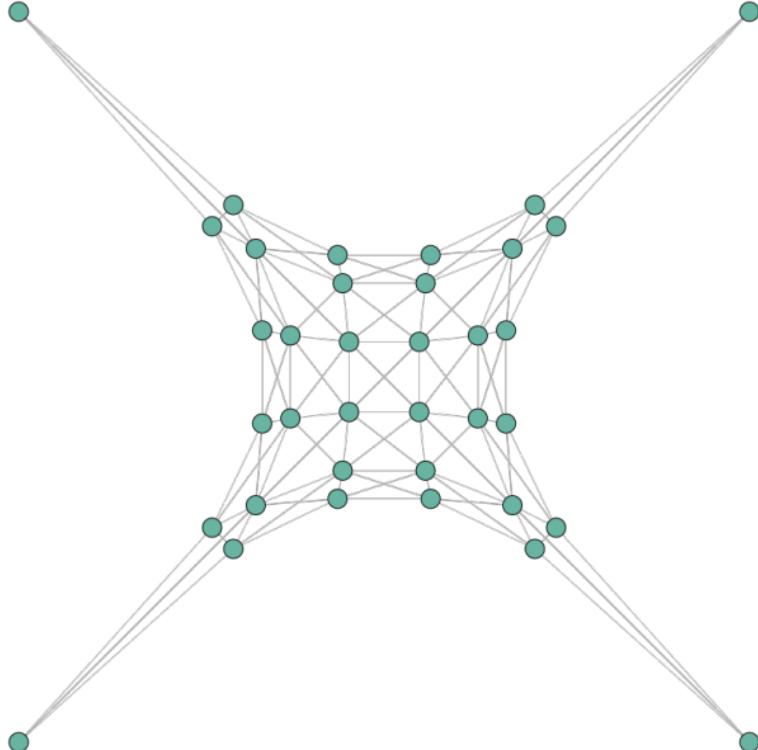


Figure 13: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

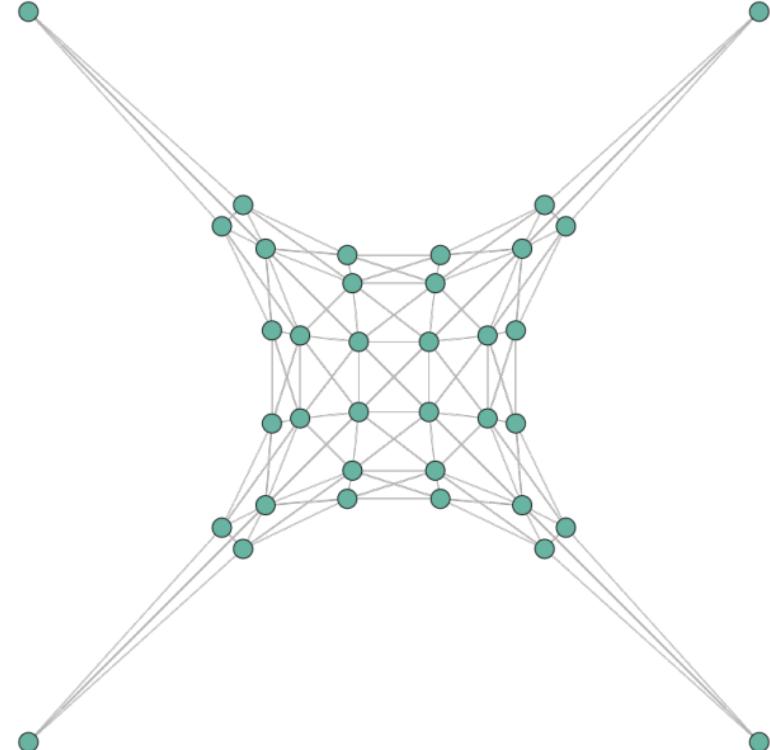


Figure 14: **Newton with projected Hessian**: 0: f without length & Constrained spring scenario: 2: Corners

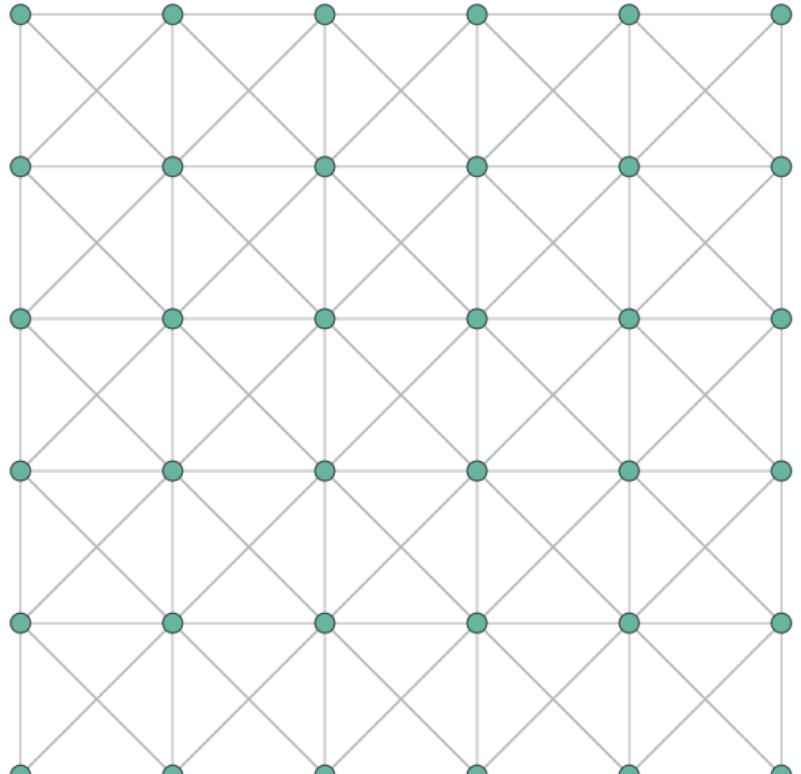


Figure 15: **Gauss Newton**: 1: f with length & Constrained spring scenario: 2: Corners

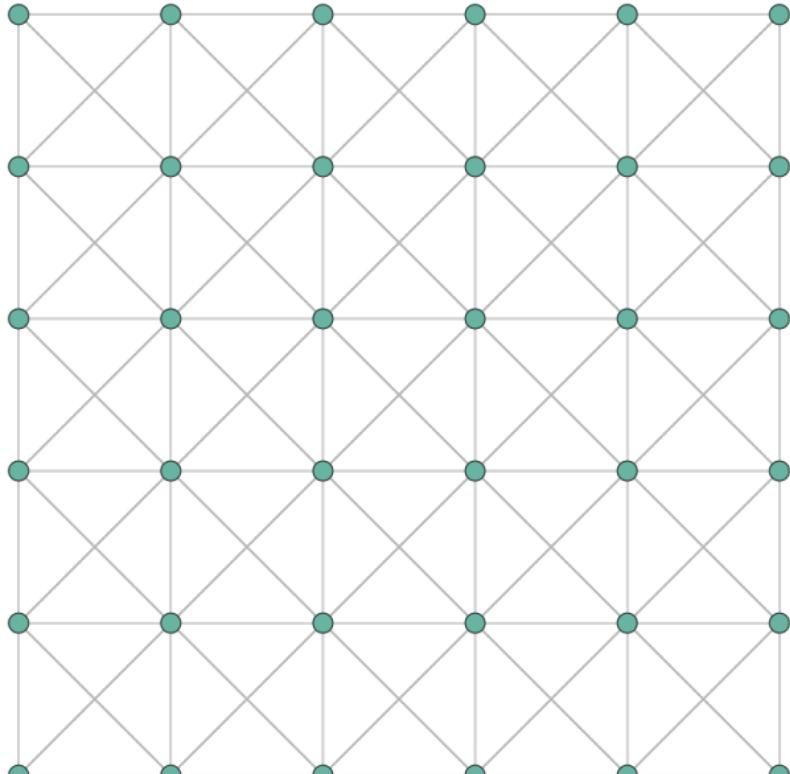


Figure 16: **Gradient Descent**: 1: f with length & Constrained spring scenario: 2: Corners

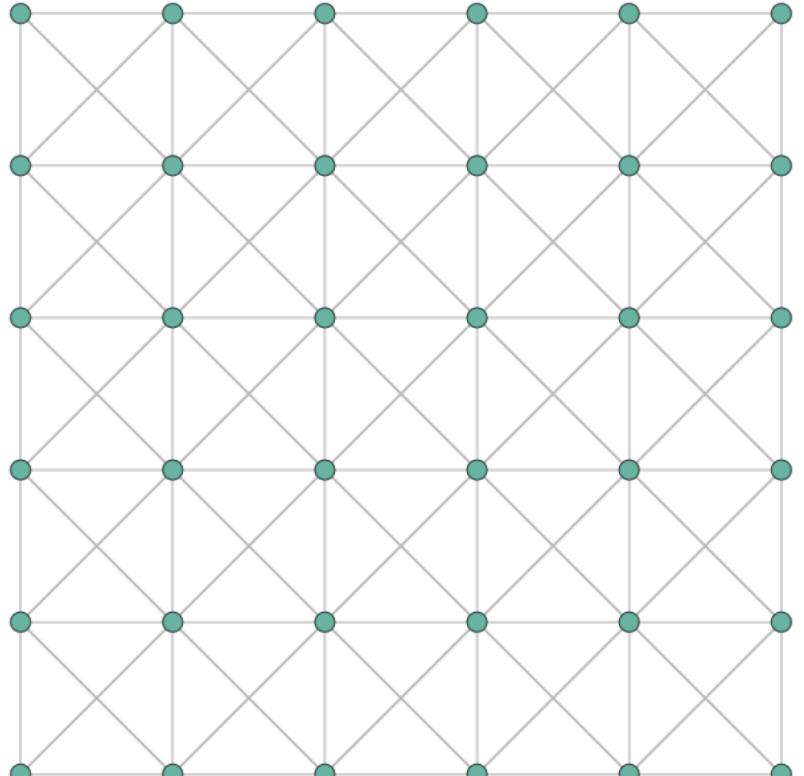


Figure 17: **Gradient Descent**: 1: f with length & Constrained spring scenario: 2: Corners

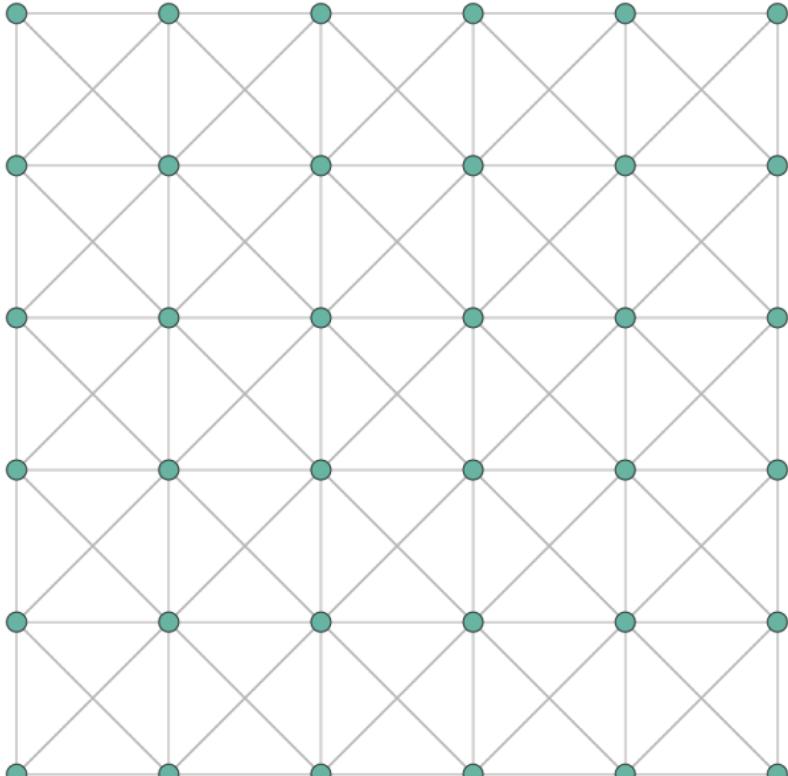


Figure 18: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

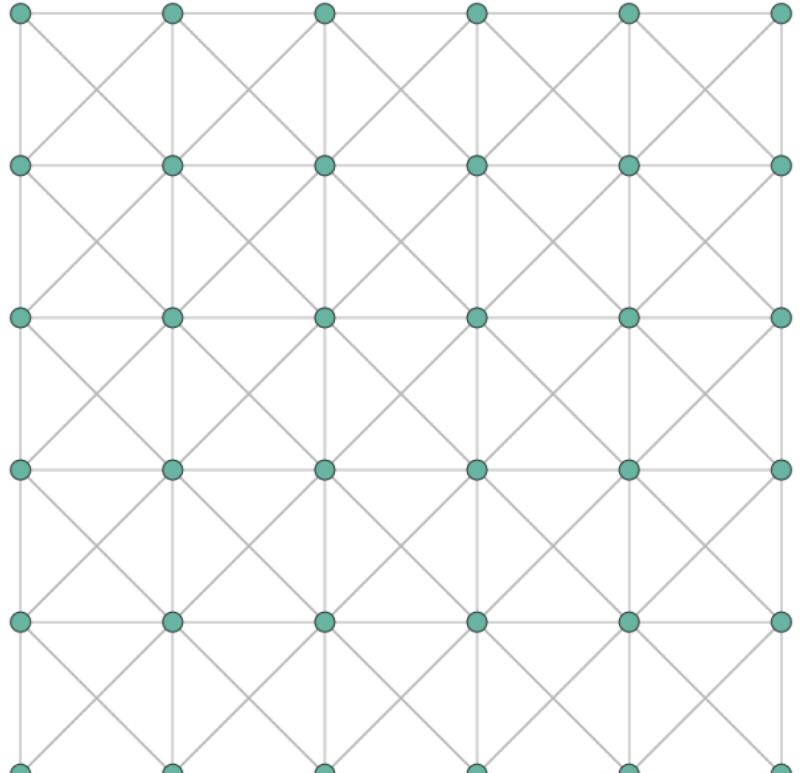


Figure 19: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

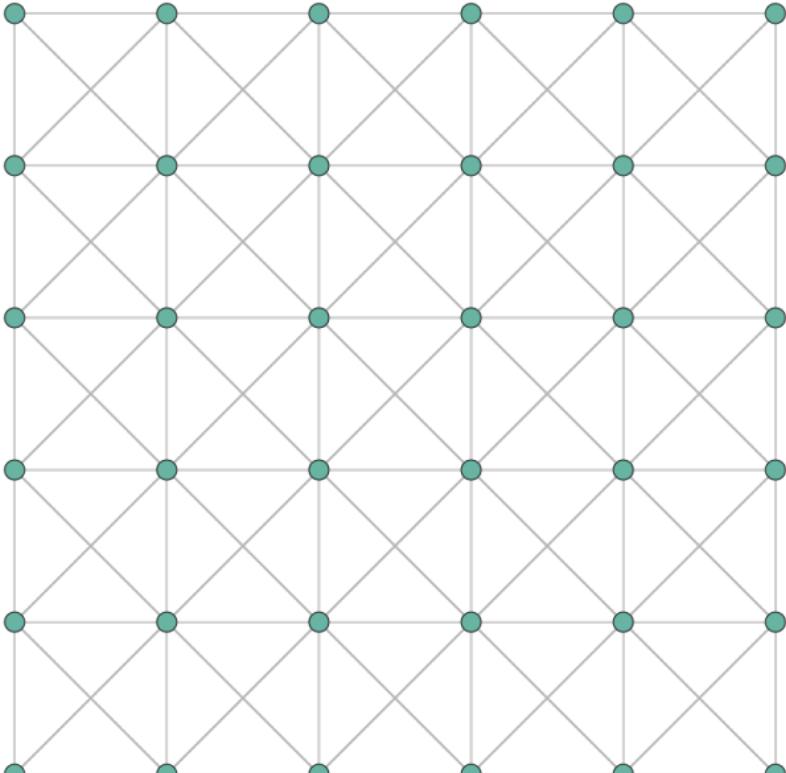


Figure 20: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

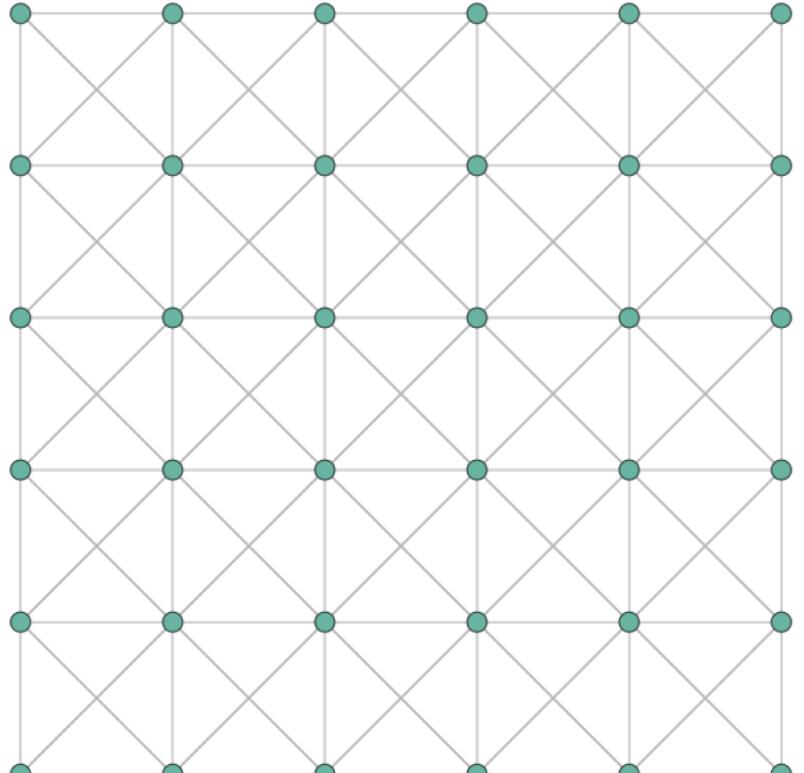


Figure 21: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

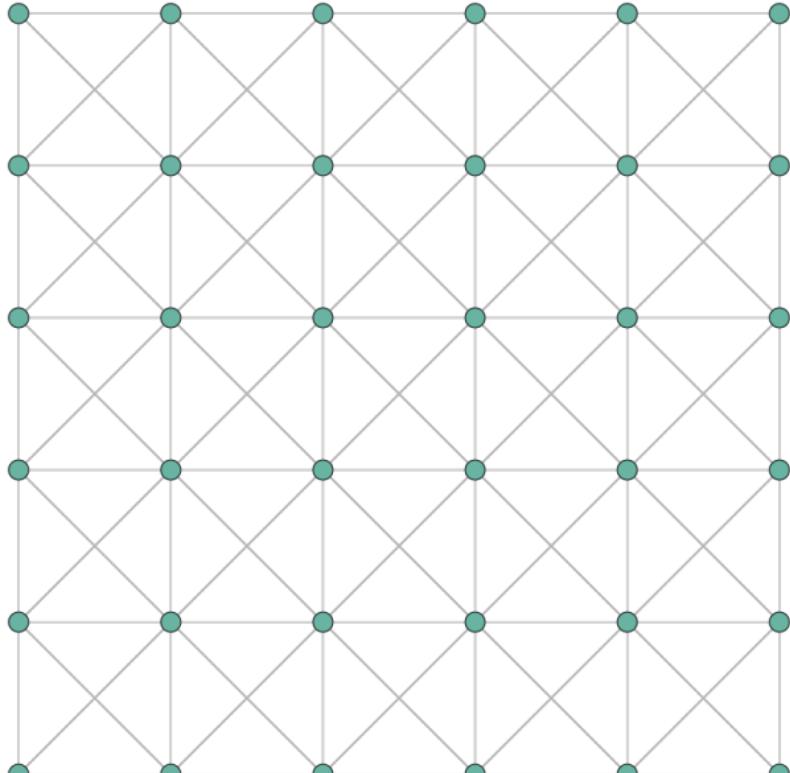


Figure 22: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

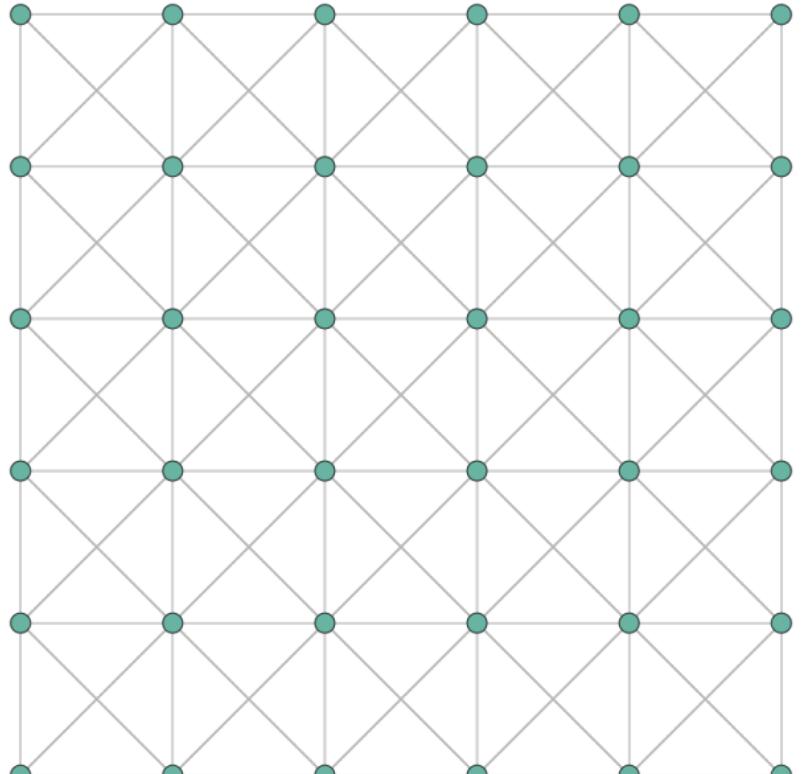


Figure 23: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

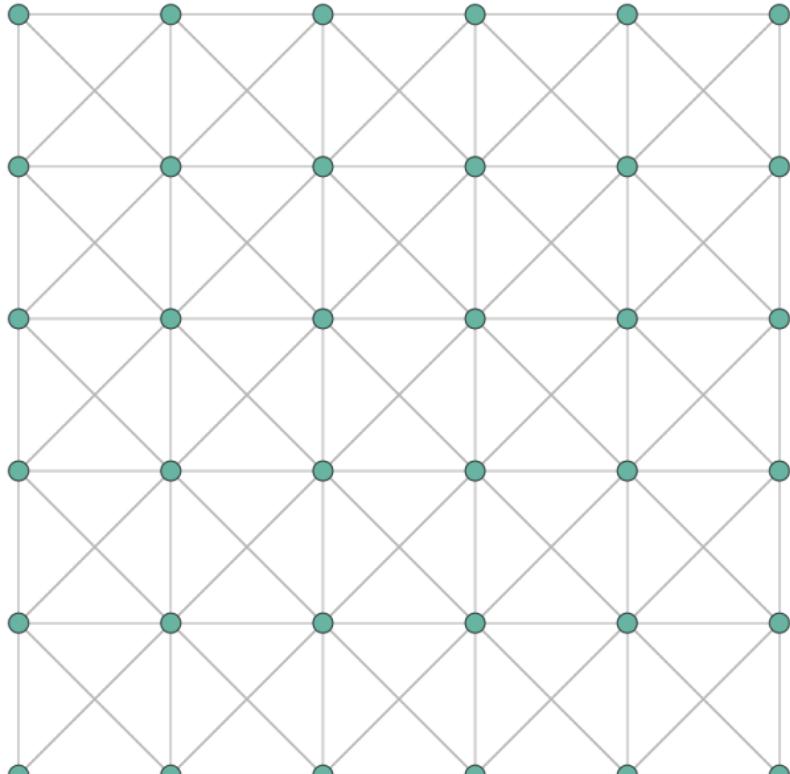


Figure 24: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

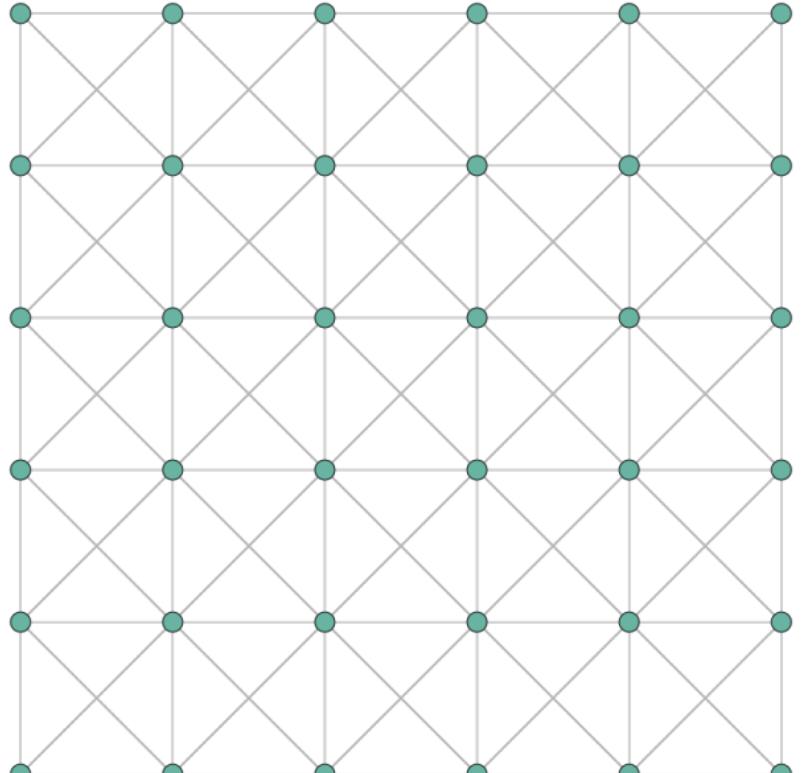


Figure 25: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

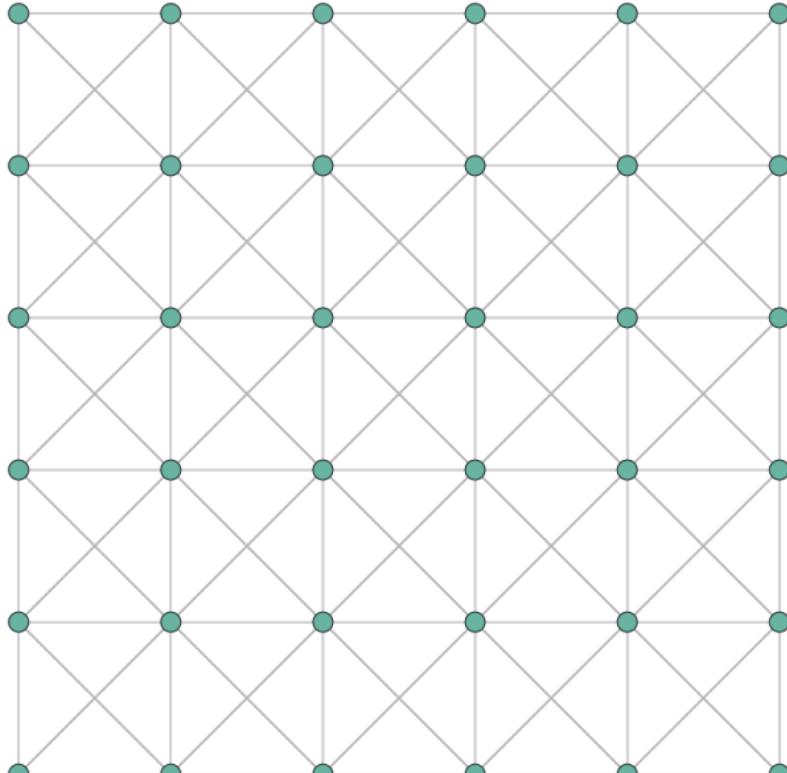


Figure 26: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

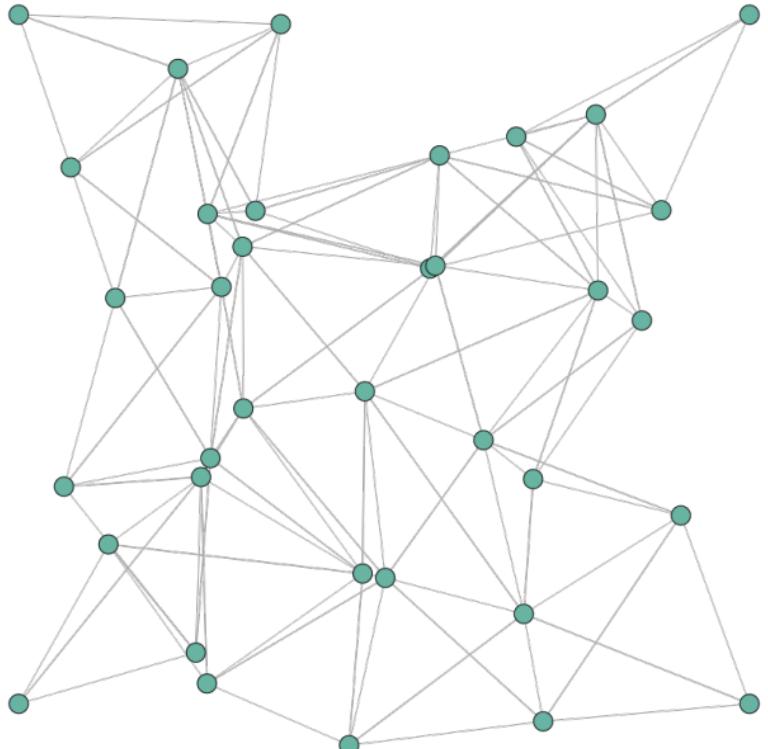


Figure 27: **Newton with projected Hessian:**
1: f with length & Constrained spring scenario:
2: Corners

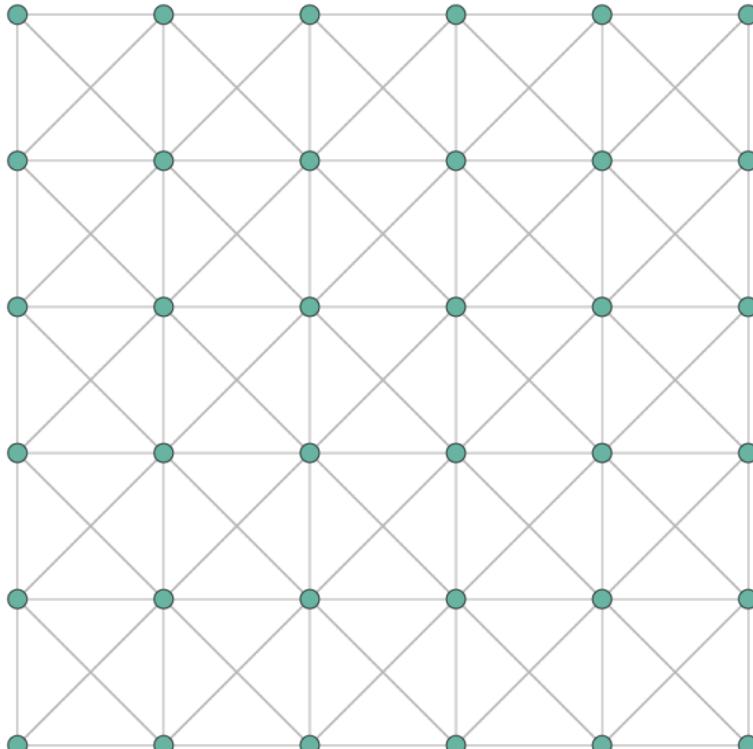


Figure 28: **Standard Newton:** 2: f with length
with positive local hessian & Constrained spring
scenario: 2: Corners

Grid size: 10

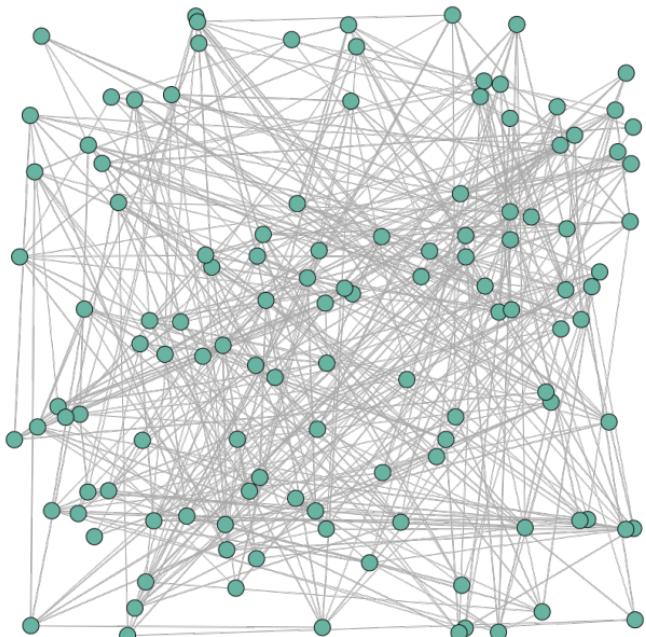


Figure 29: **Gauss Newton:** 0: f without length & Constrained spring scenario: 2: Corners – (As Gauss Newton executable uses the Newton Methods without hessian modification, and in this case, the LLT factorization fails, which makes marks the end of the method like all other **Gauss Newton without length**)

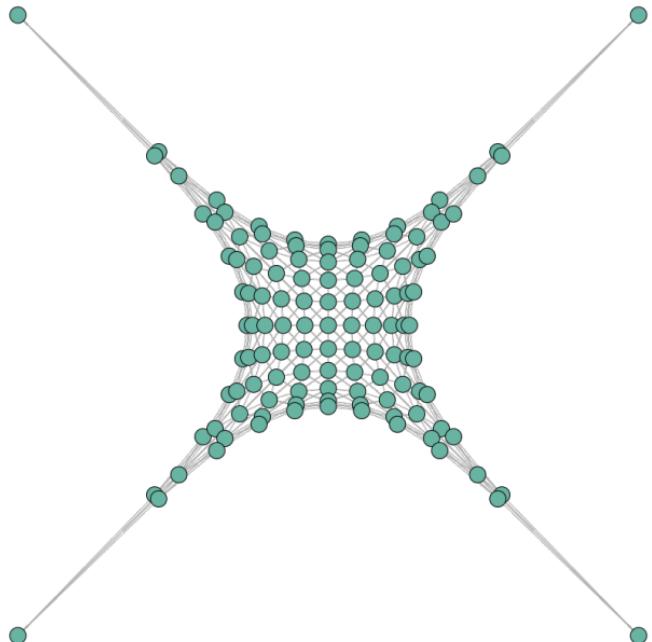


Figure 30: **Standard Newton:** 0: f without length & Constrained spring scenario: 2: Corners

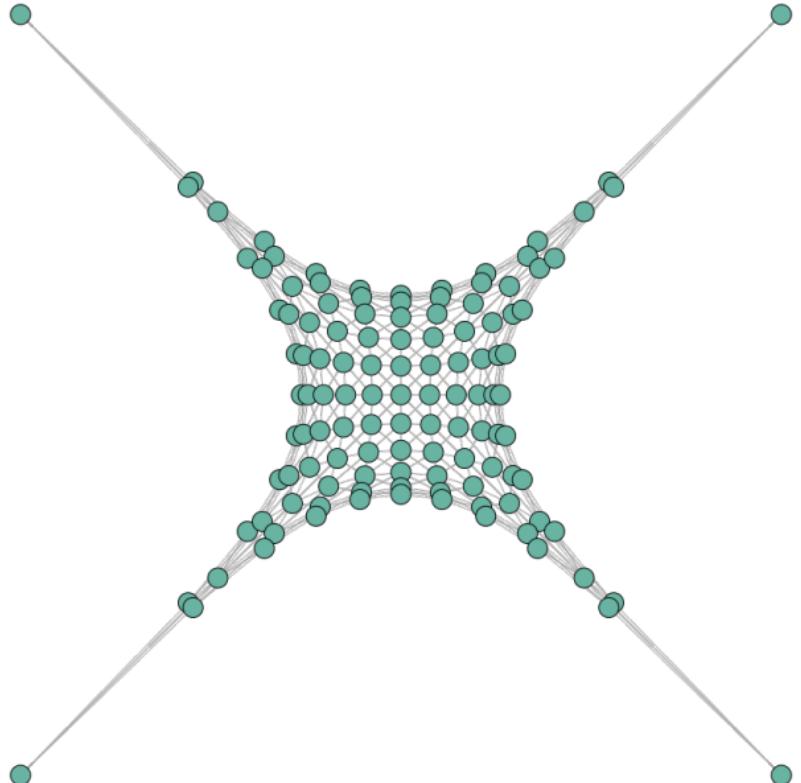


Figure 31: **Gradient Descent**: 0: f without length & Constrained spring scenario: 1: Sides

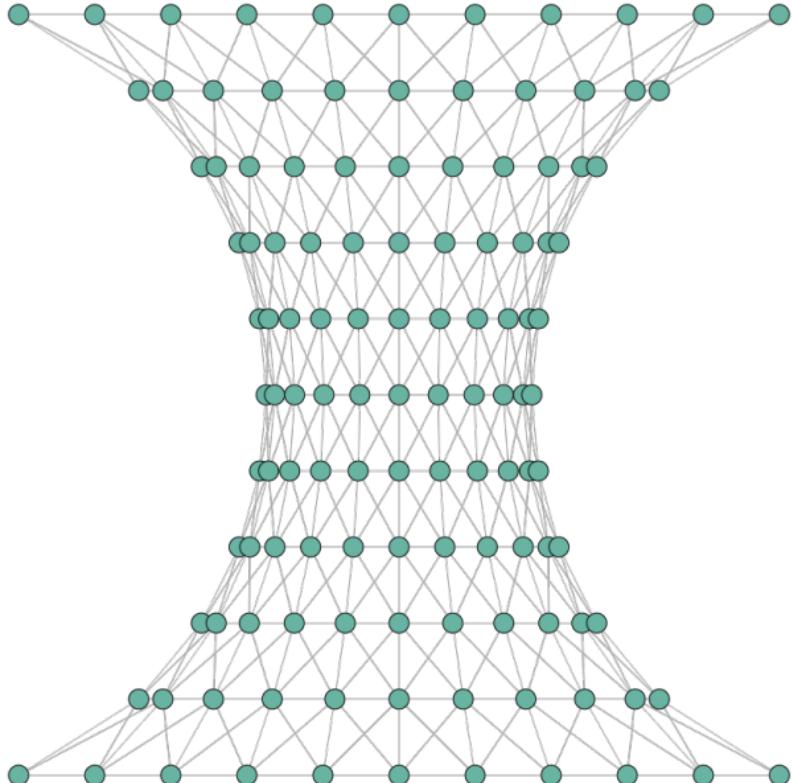


Figure 32: **Gradient Descent**: 0: f without length & Constrained spring scenario: 1: Sides

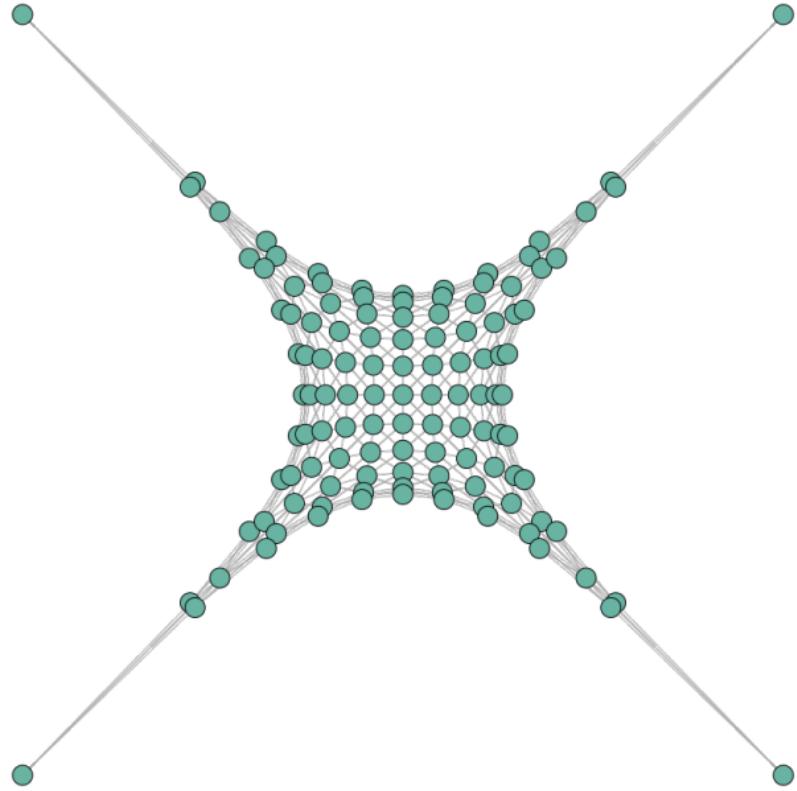


Figure 33: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

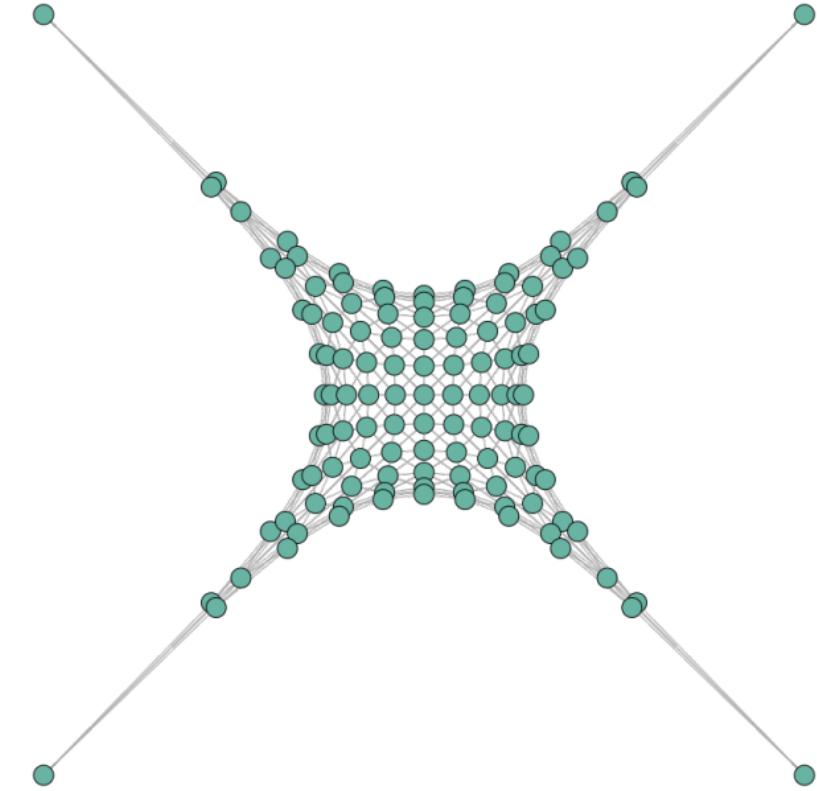


Figure 34: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

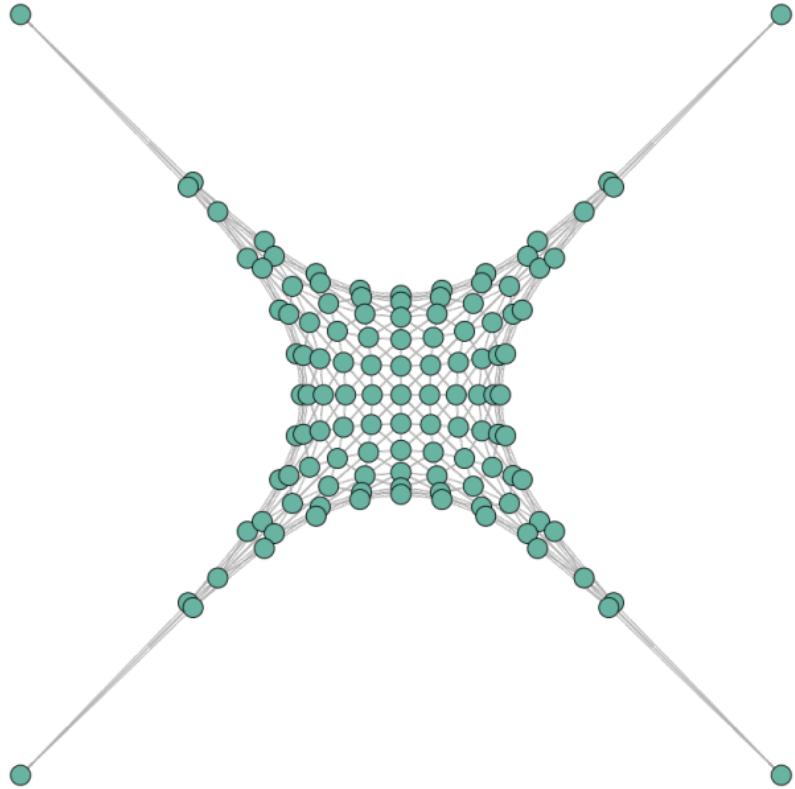


Figure 35: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

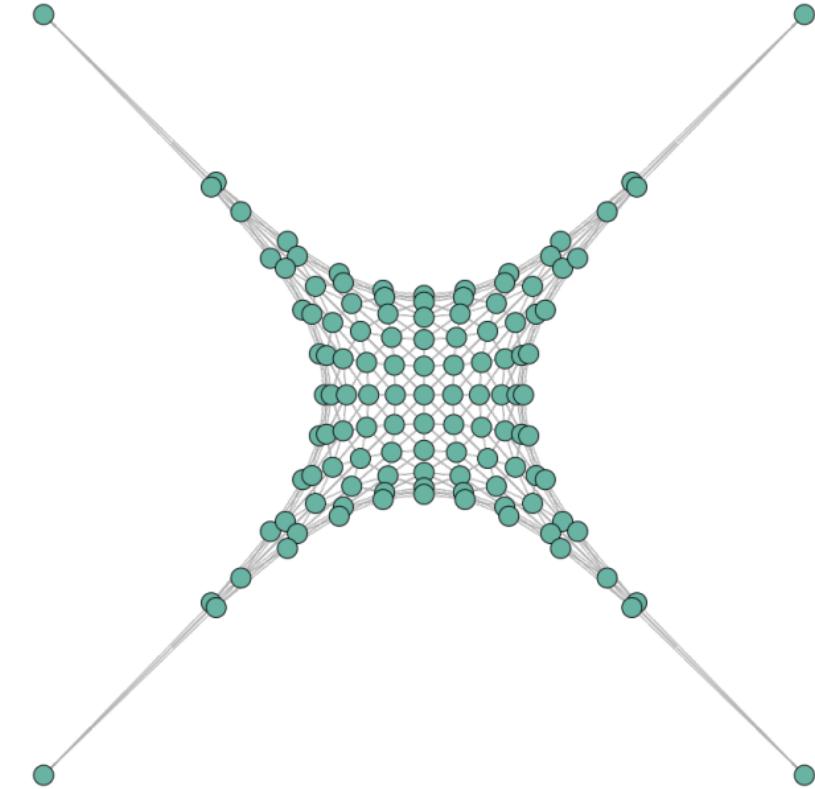


Figure 36: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

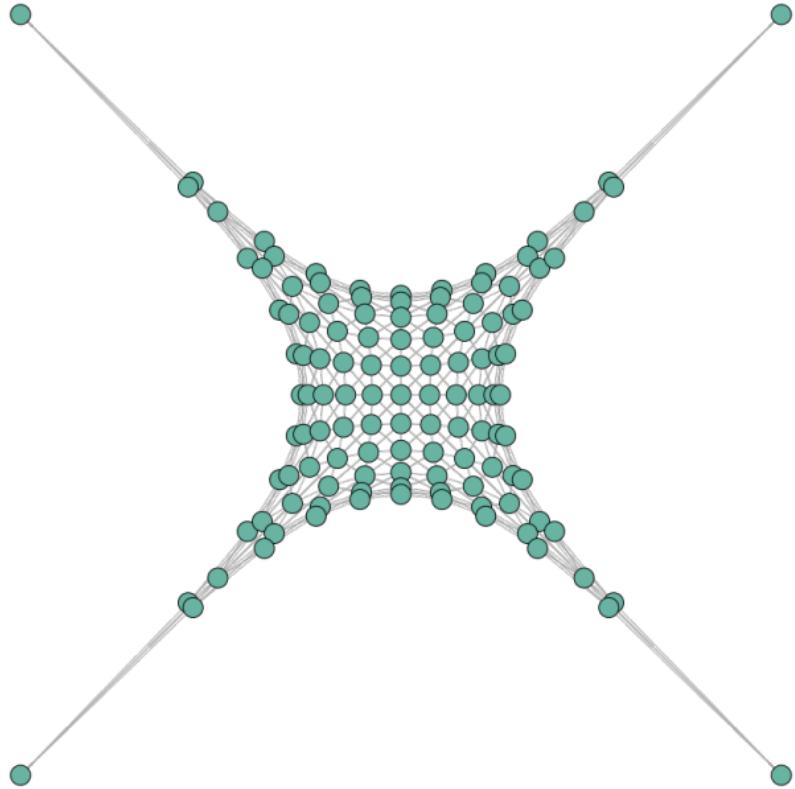


Figure 37: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

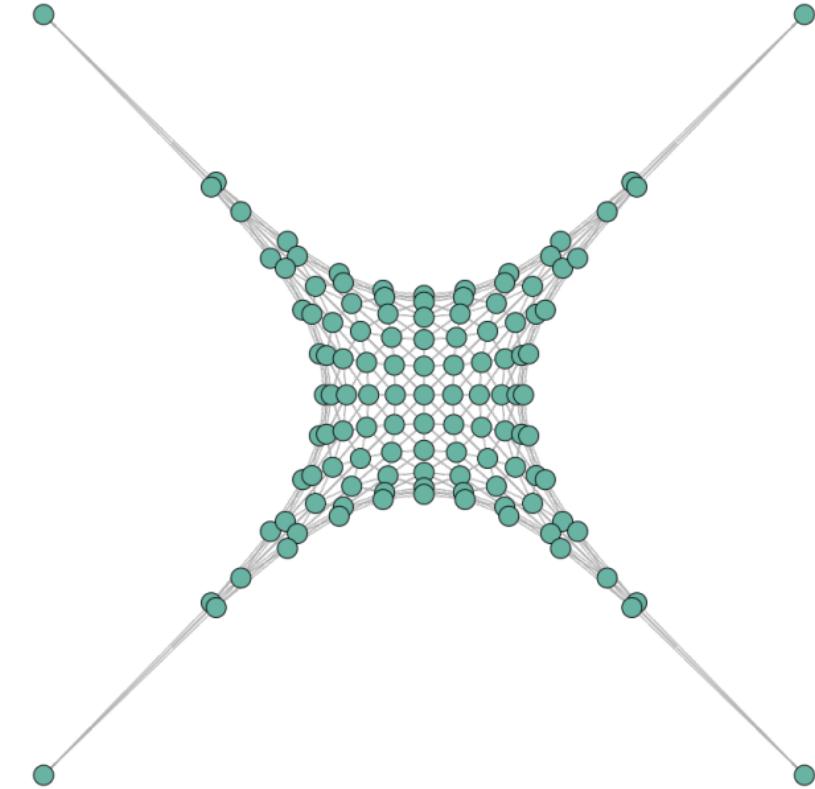


Figure 38: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

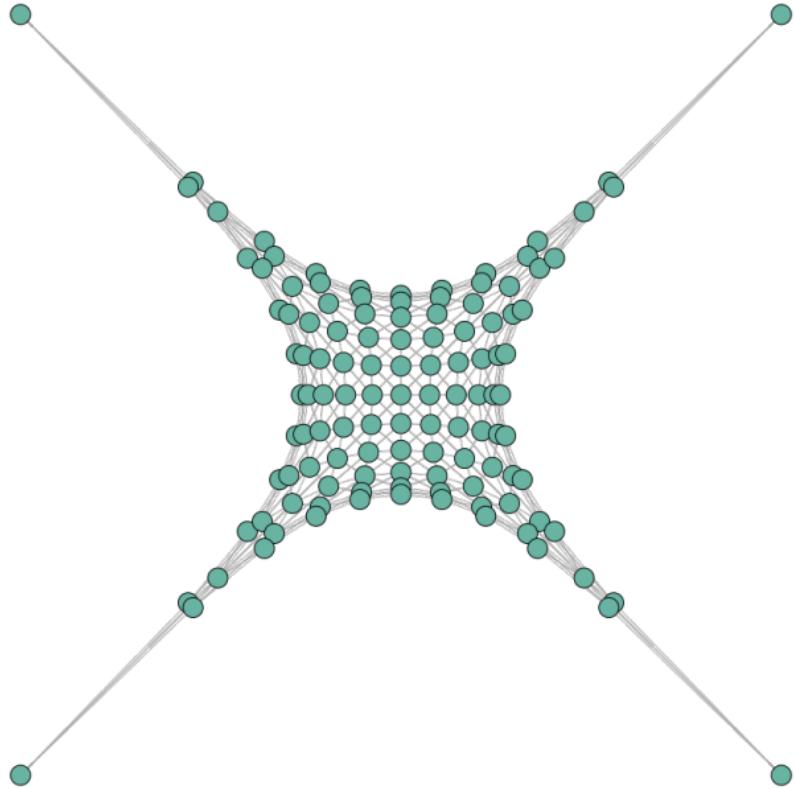


Figure 39: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

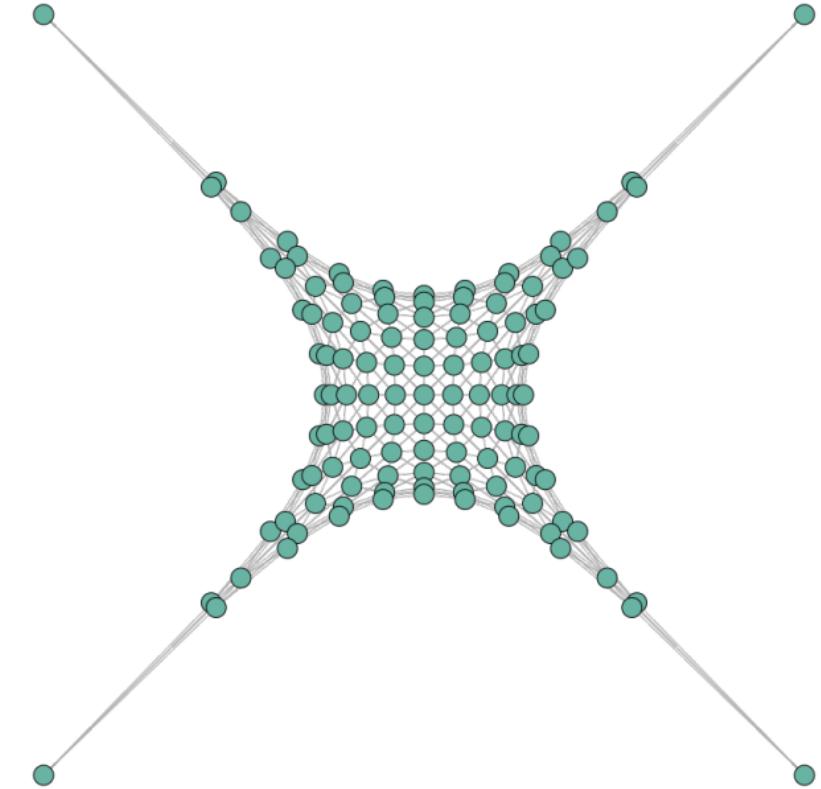


Figure 40: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

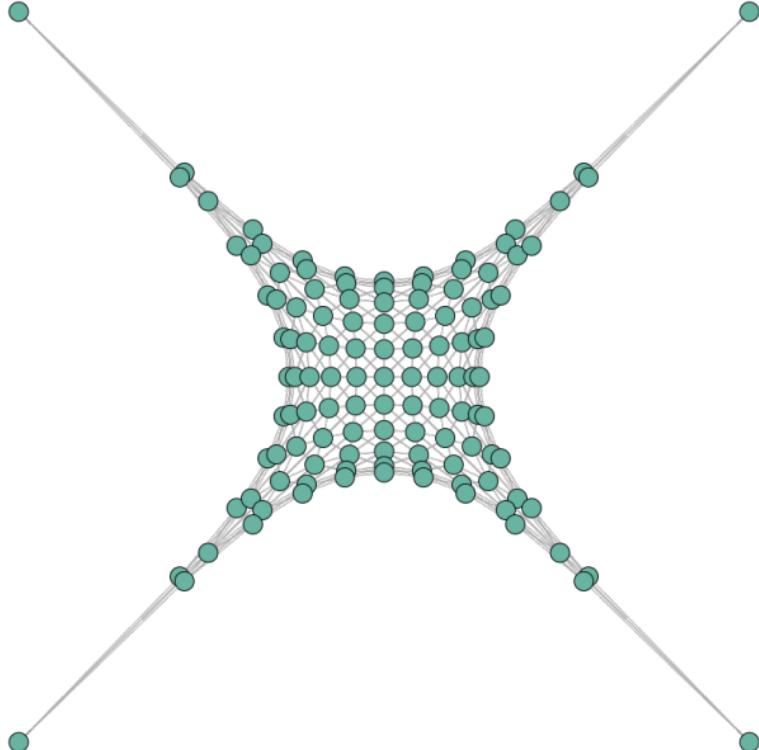


Figure 41: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

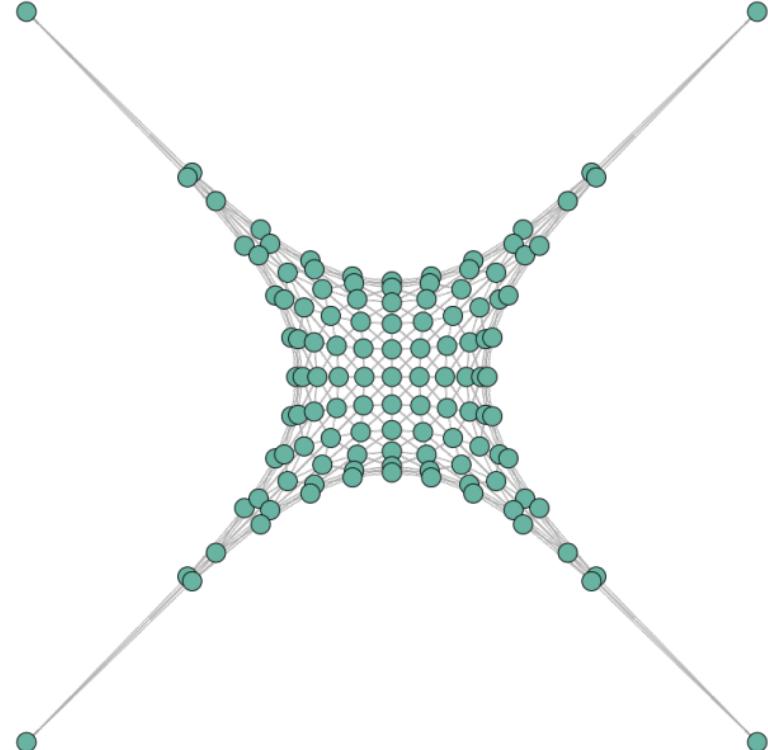


Figure 42: **Newton with projected Hessian**: 0: f without length & Constrained spring scenario: 2: Corners

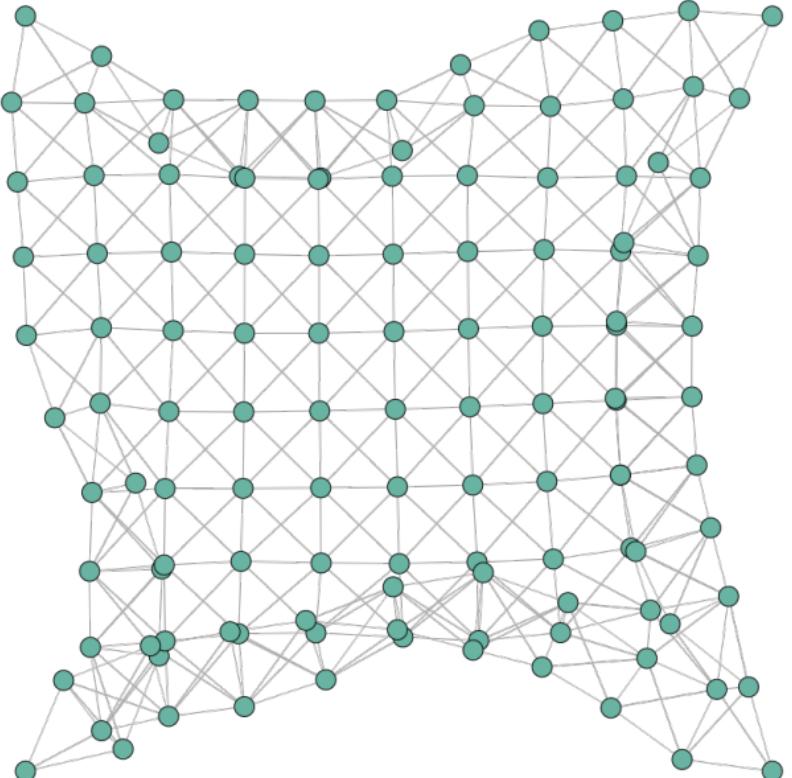


Figure 43: **Gauss Newton**: 1: f with length & Constrained spring scenario: 2: Corners

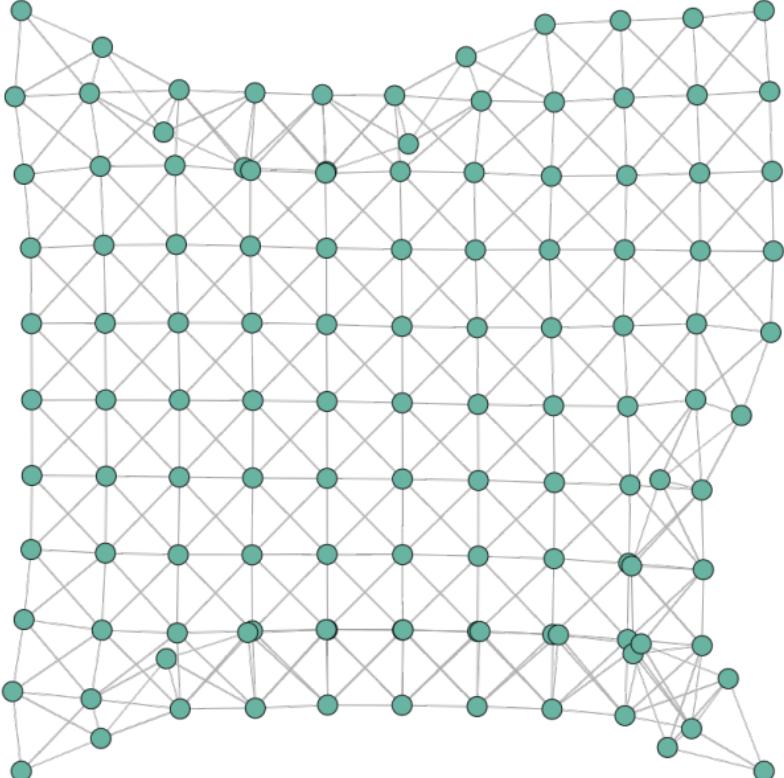


Figure 44: **Gradient Descent**: 1: f with length & Constrained spring scenario: 2: Corners

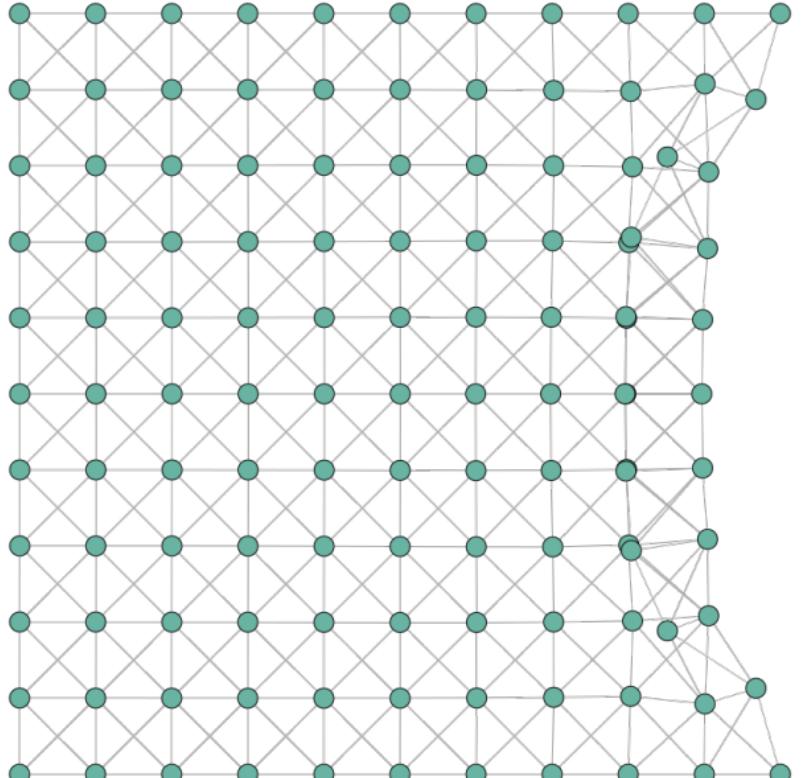


Figure 45: **Gradient Descent**: 1: f with length & Constrained spring scenario: 2: Corners

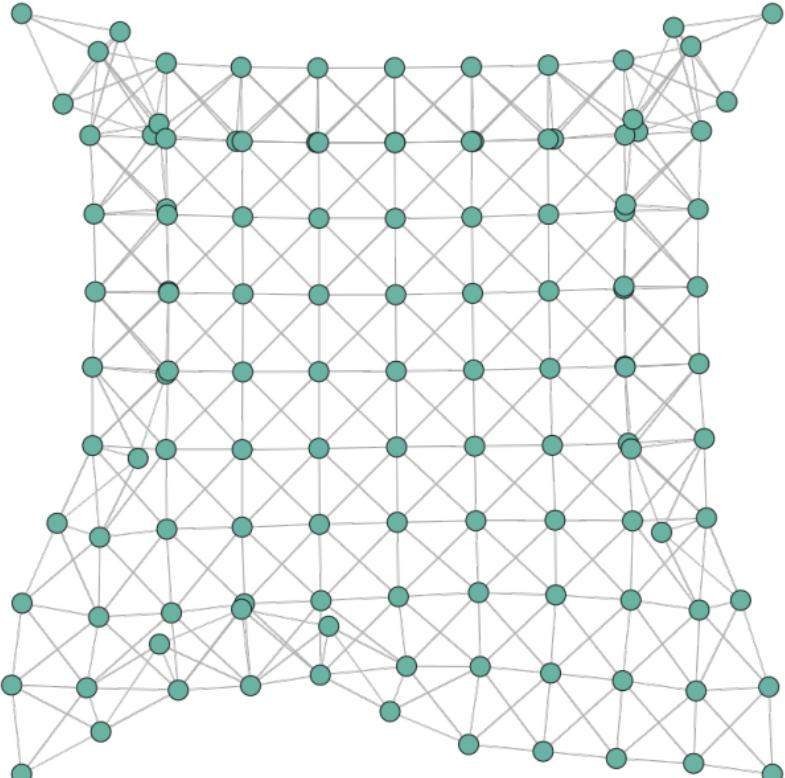


Figure 46: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

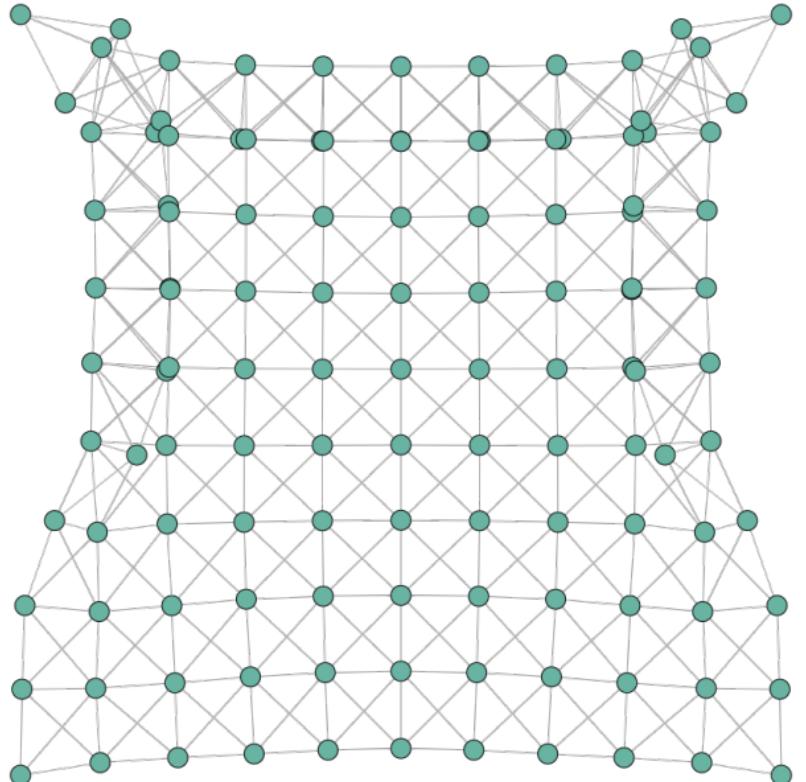


Figure 47: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

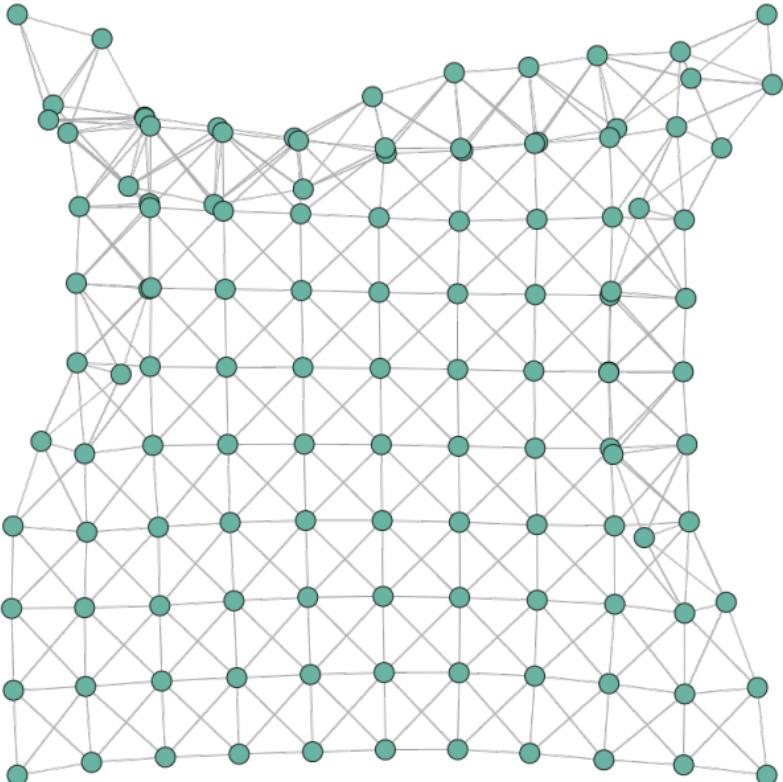


Figure 48: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

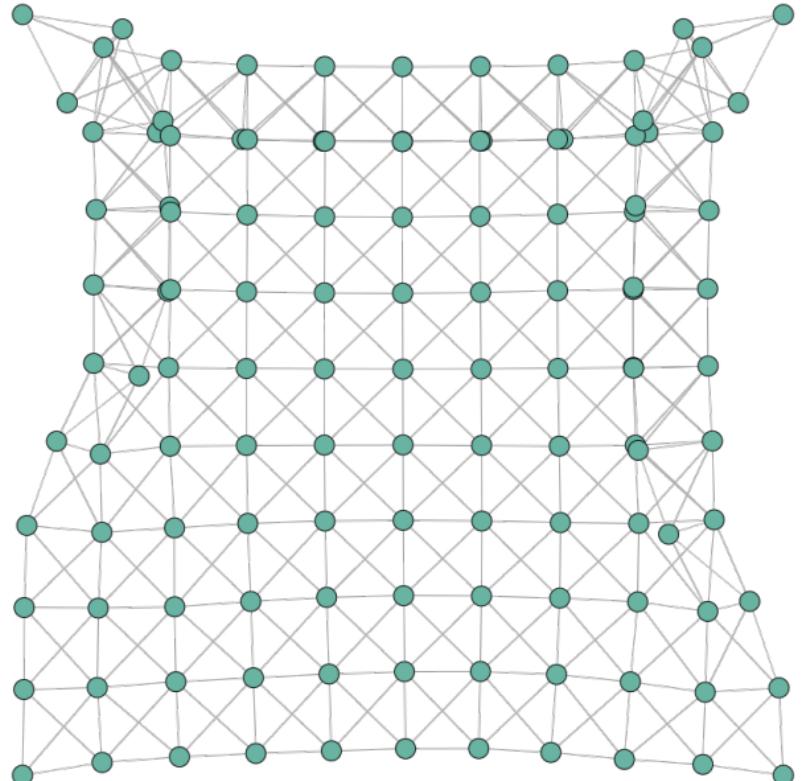


Figure 49: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

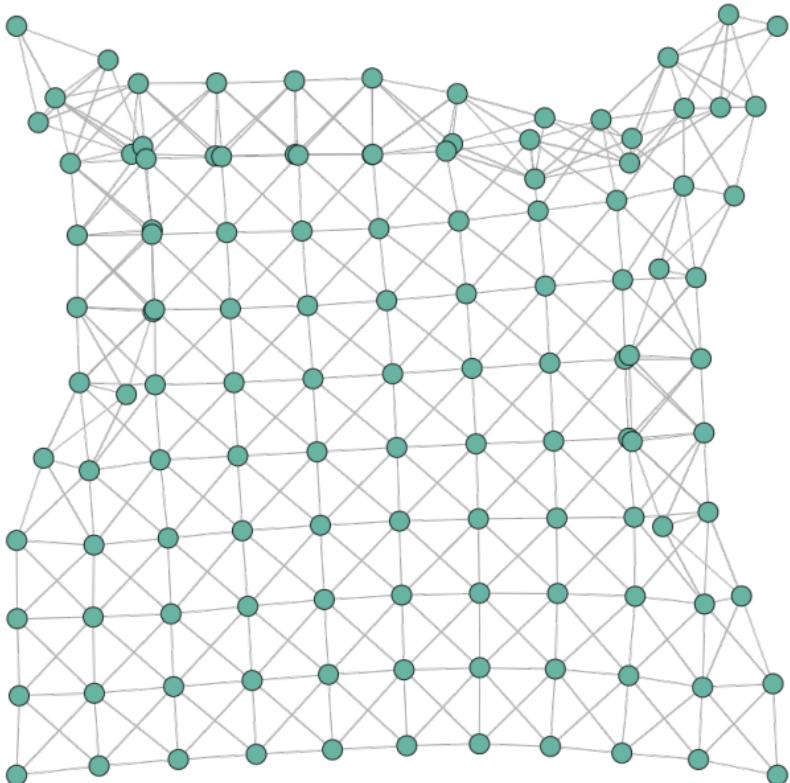


Figure 50: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

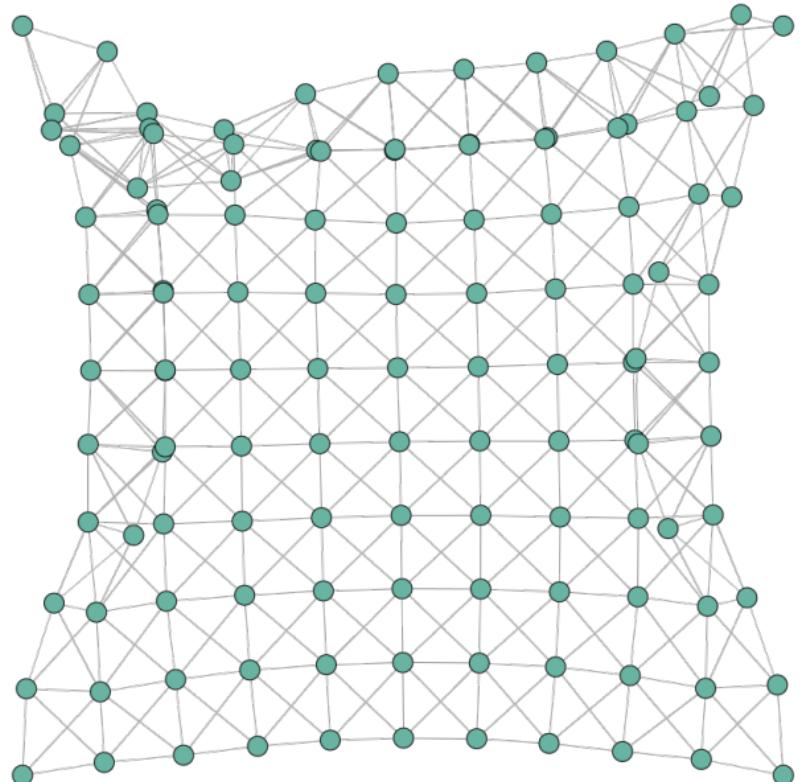


Figure 51: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

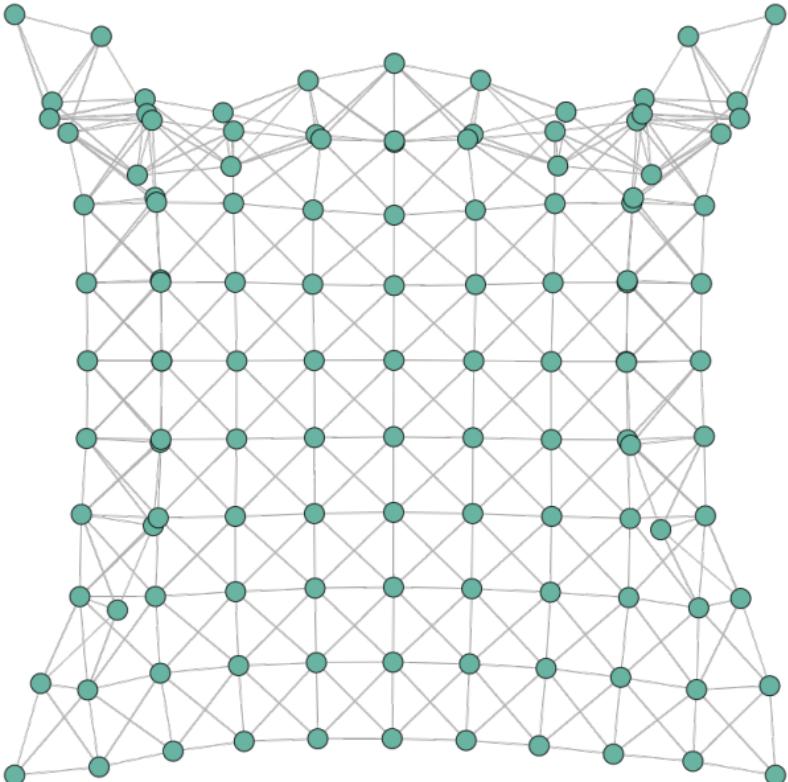


Figure 52: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

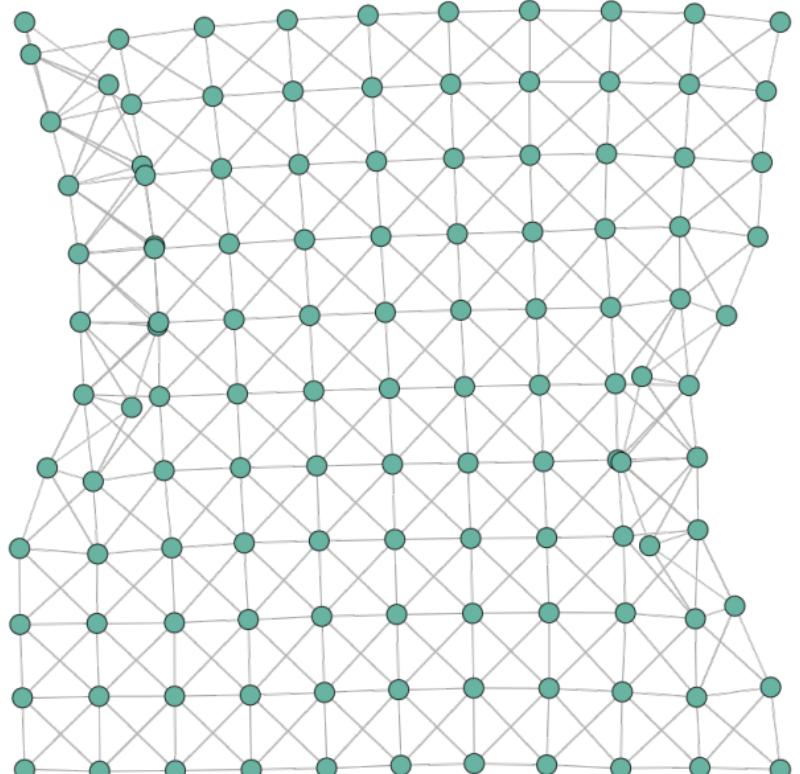


Figure 53: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

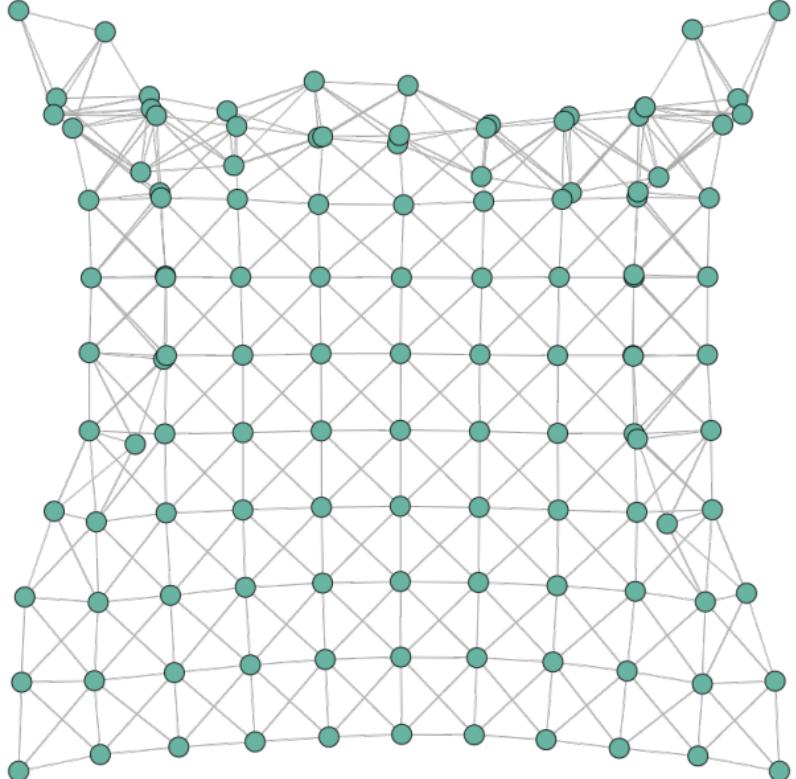


Figure 54: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners



Figure 55: **Newton with projected Hessian:**
1: f with length & Constrained spring scenario:
2: Corners

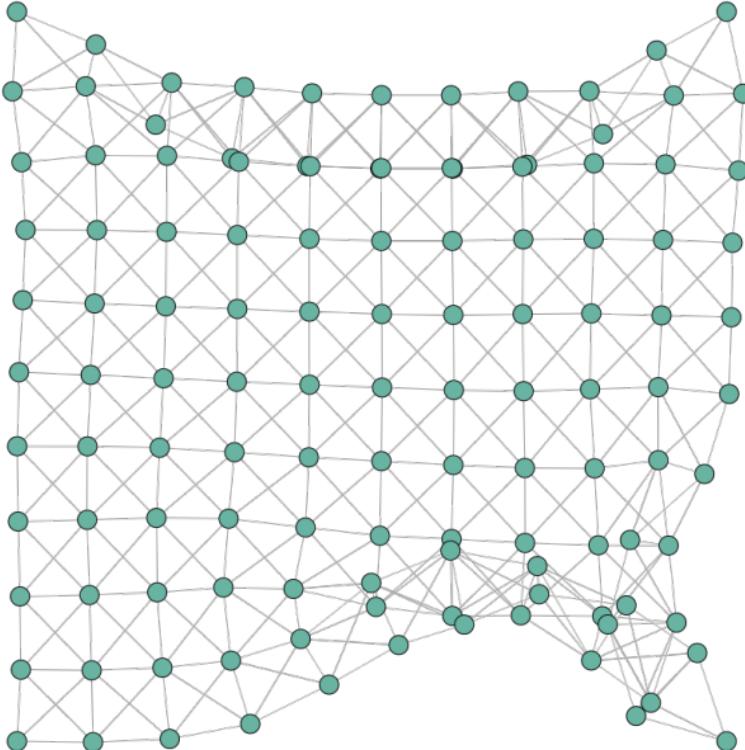


Figure 56: **Standard Newton:** 2: f with length
with positive local hessian & Constrained spring
scenario: 2: Corners

Grid size: 20

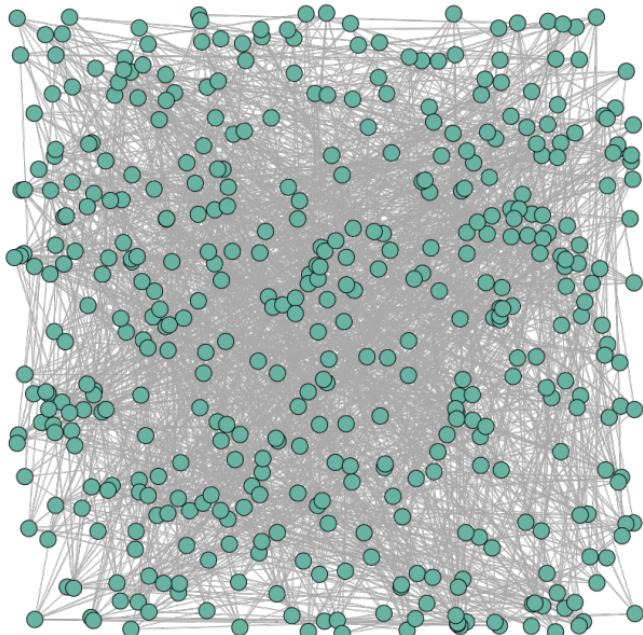


Figure 57: **Gauss Newton:** 0: f without length & Constrained spring scenario: 2: Corners – (As Gauss Newton executable uses the Newton Methods without hessian modification, and in this case, the LLT factorization fails, which makes marks the end of the method like all other **Gauss Newton without length**)

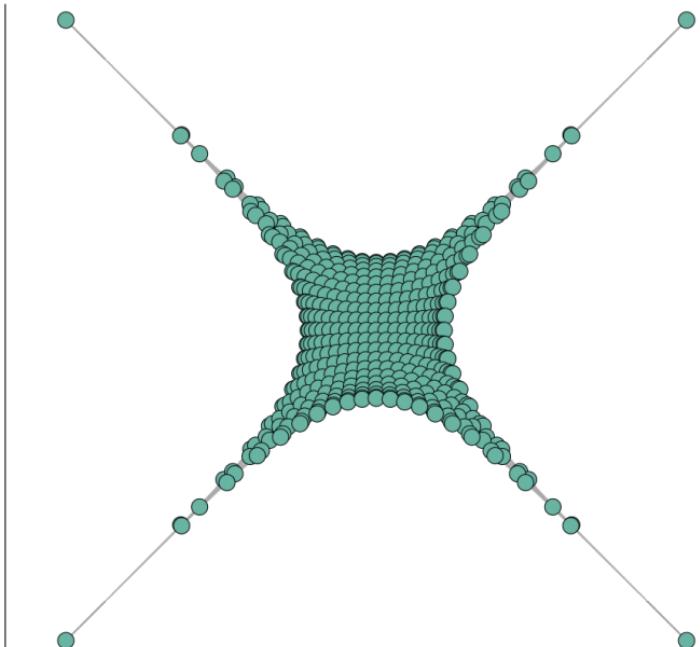


Figure 58: **Standard Newton:** 0: f without length & Constrained spring scenario: 2: Corners

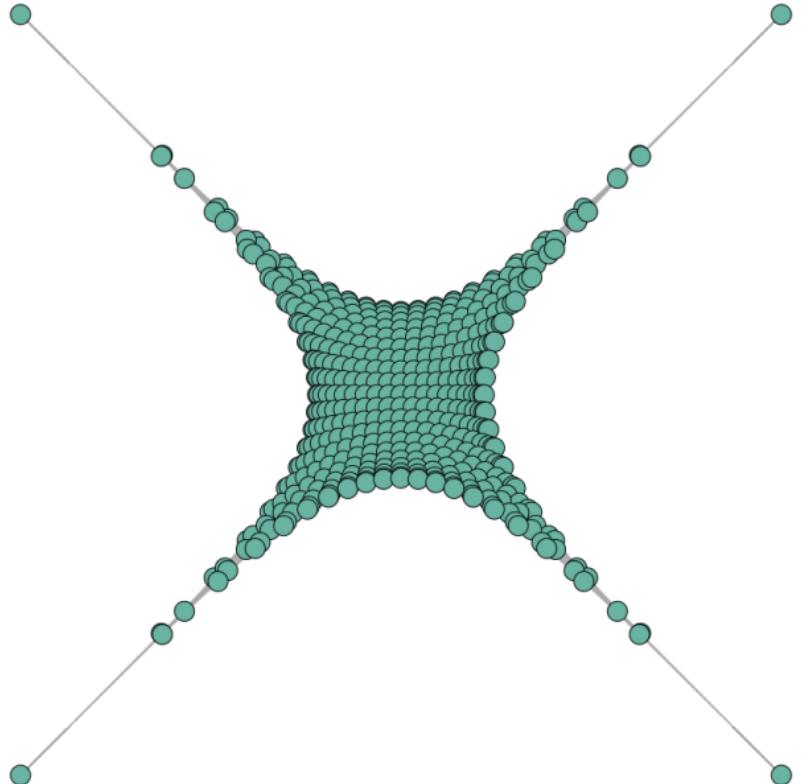


Figure 59: **Gradient Descent**: 0: f without length & Constrained spring scenario: 1: Sides

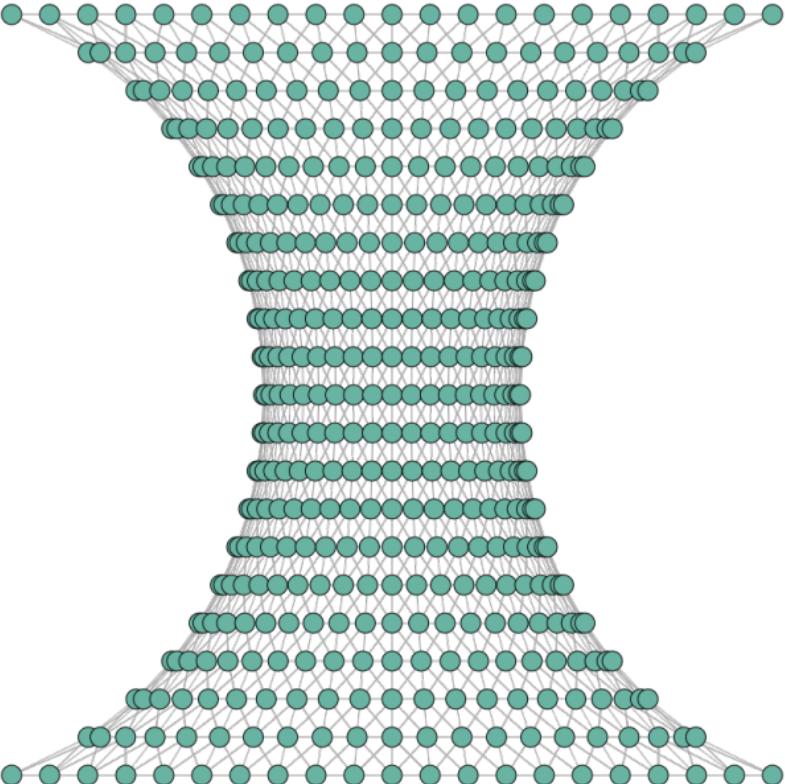


Figure 60: **Gradient Descent**: 0: f without length & Constrained spring scenario: 1: Sides

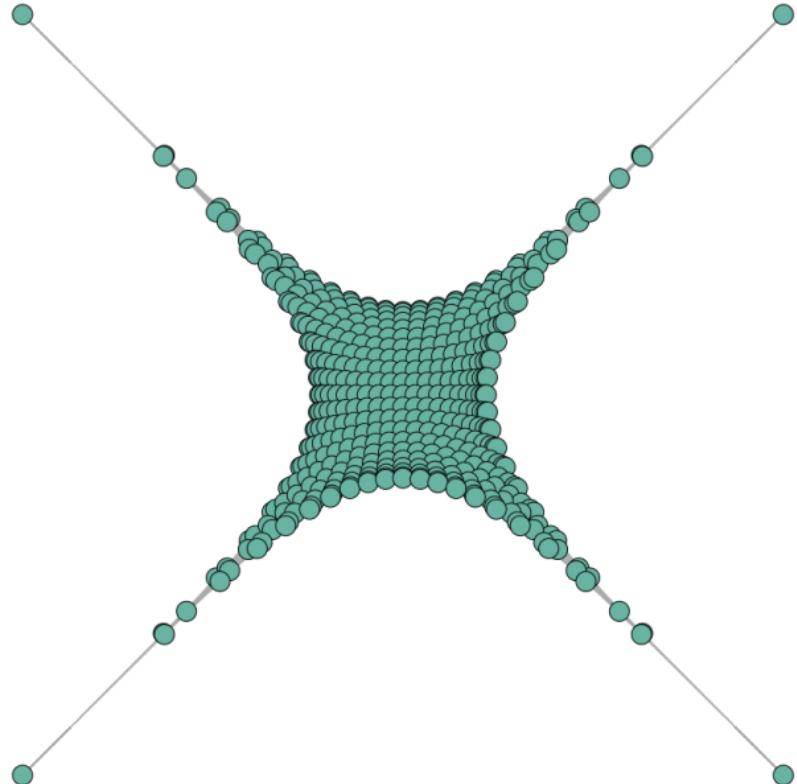


Figure 61: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

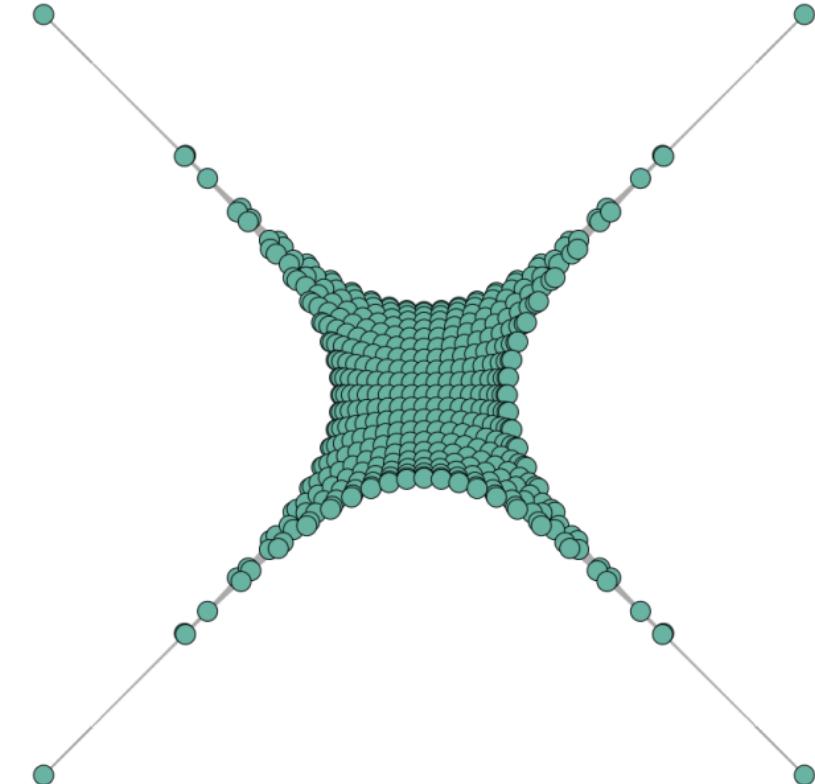


Figure 62: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

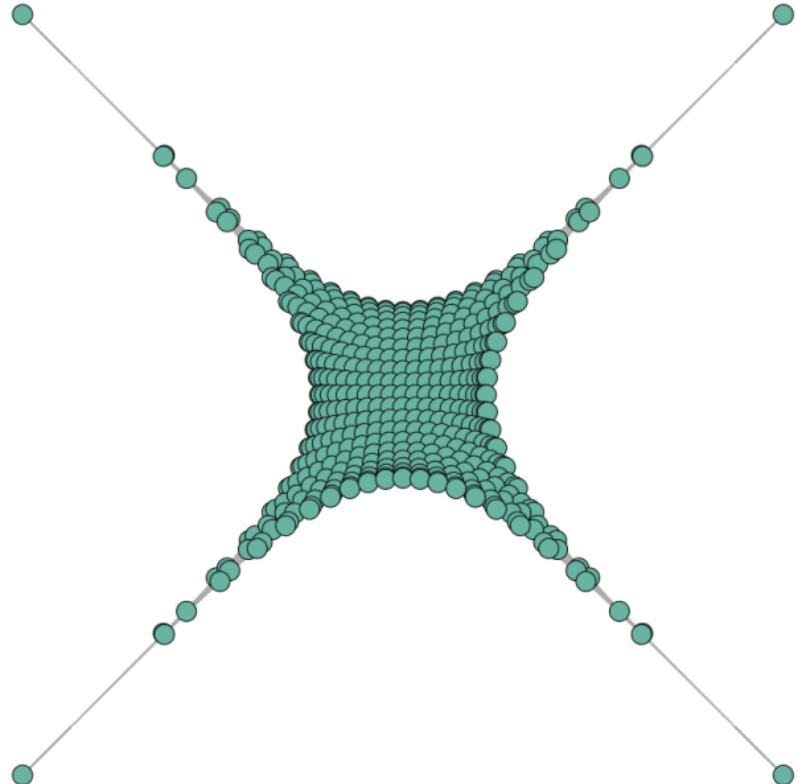


Figure 63: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

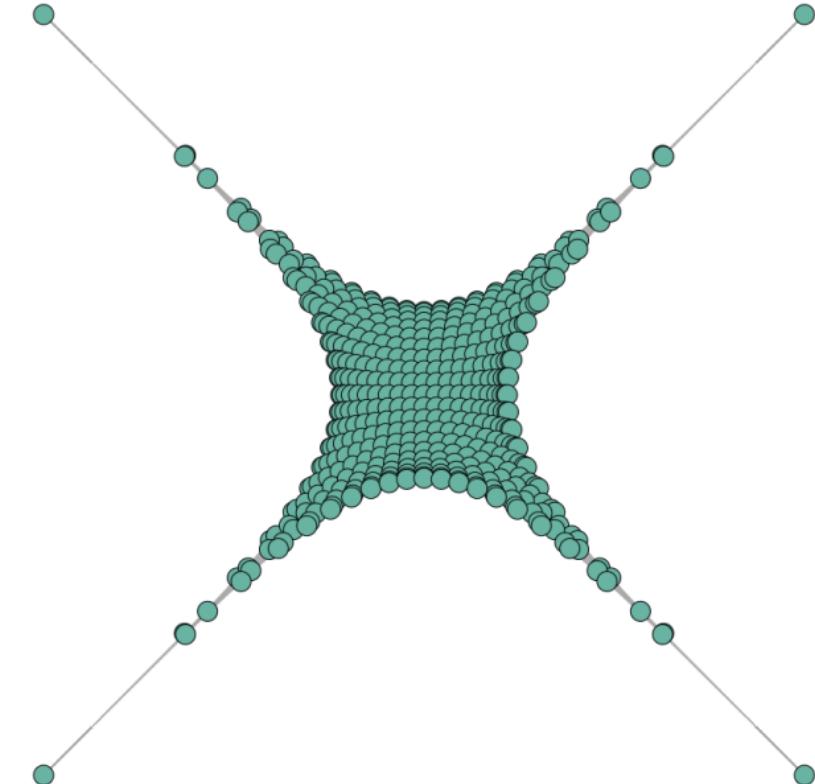


Figure 64: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

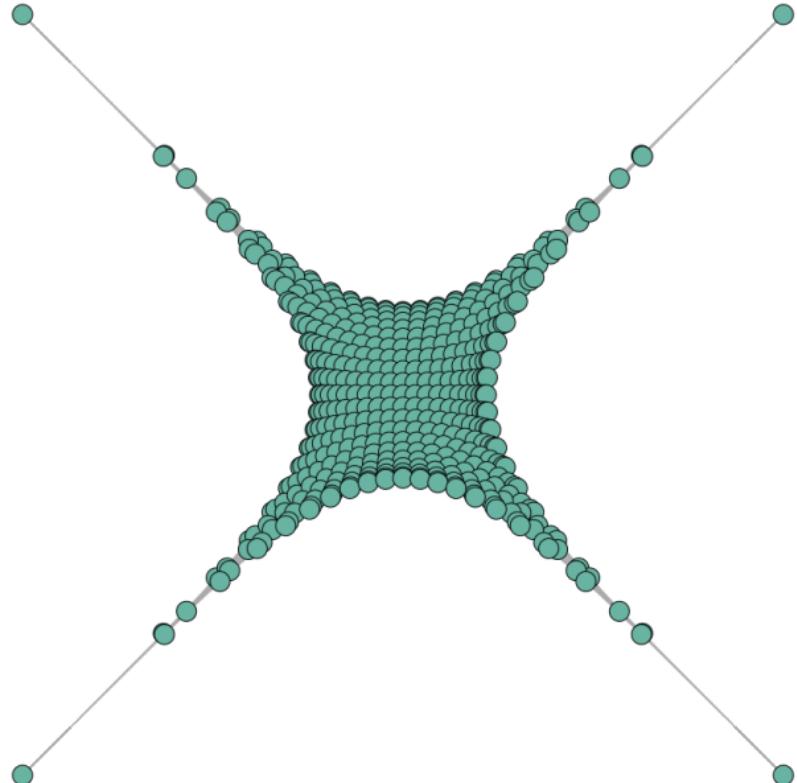


Figure 65: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

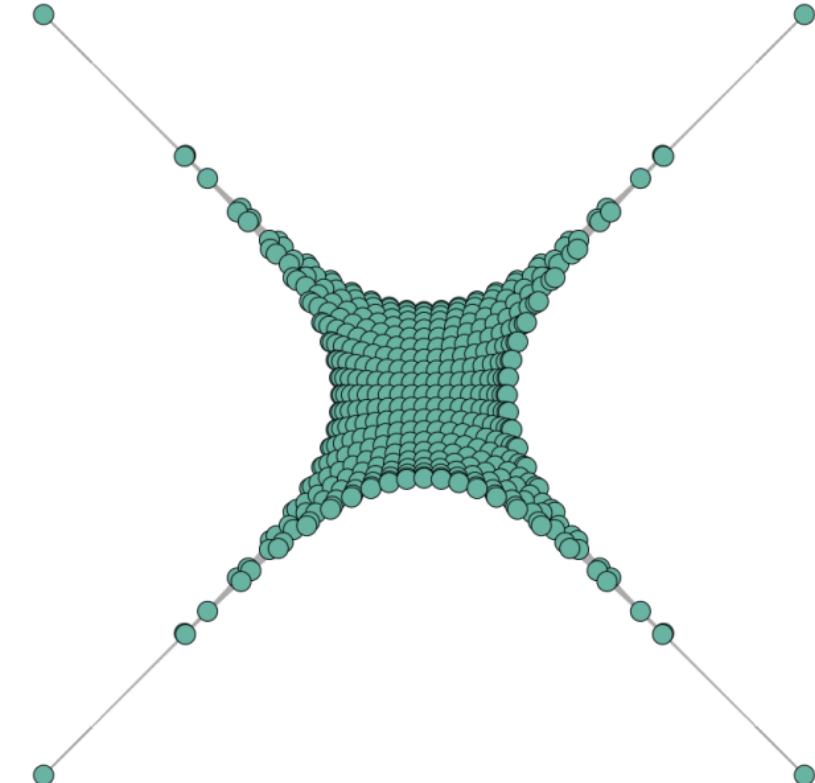


Figure 66: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

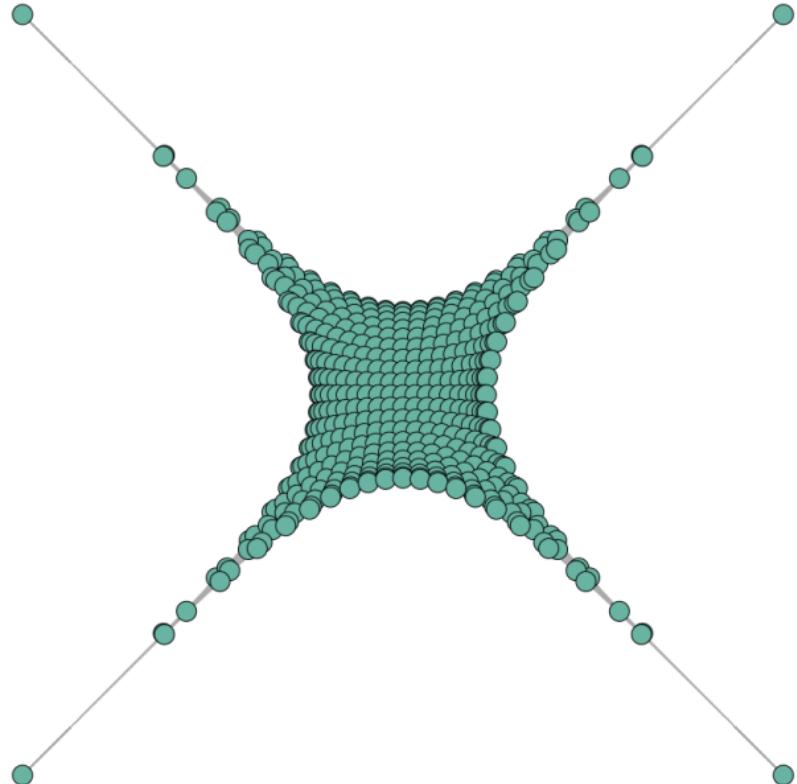


Figure 67: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

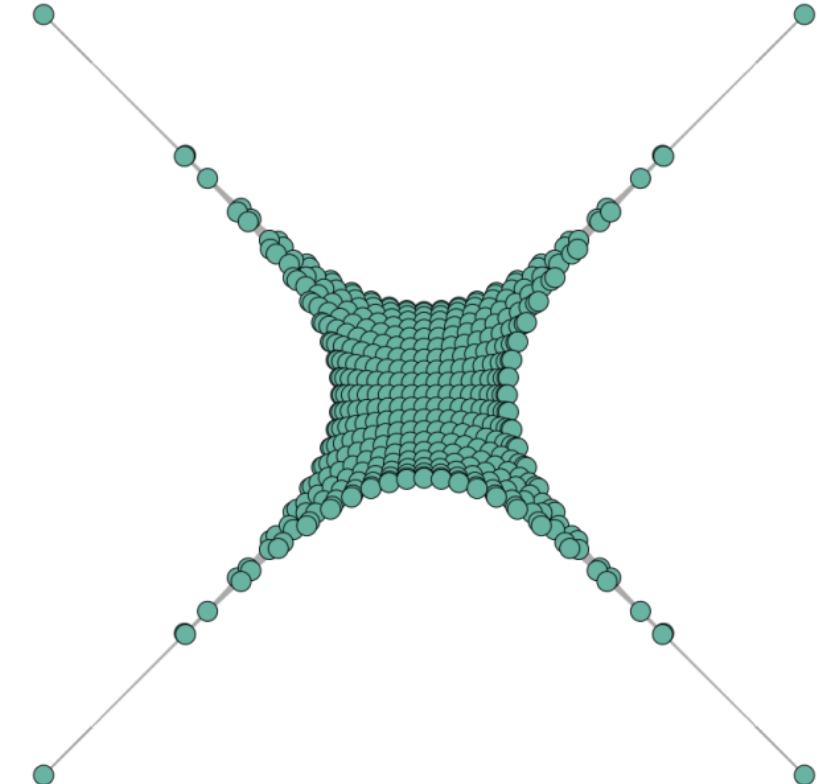


Figure 68: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

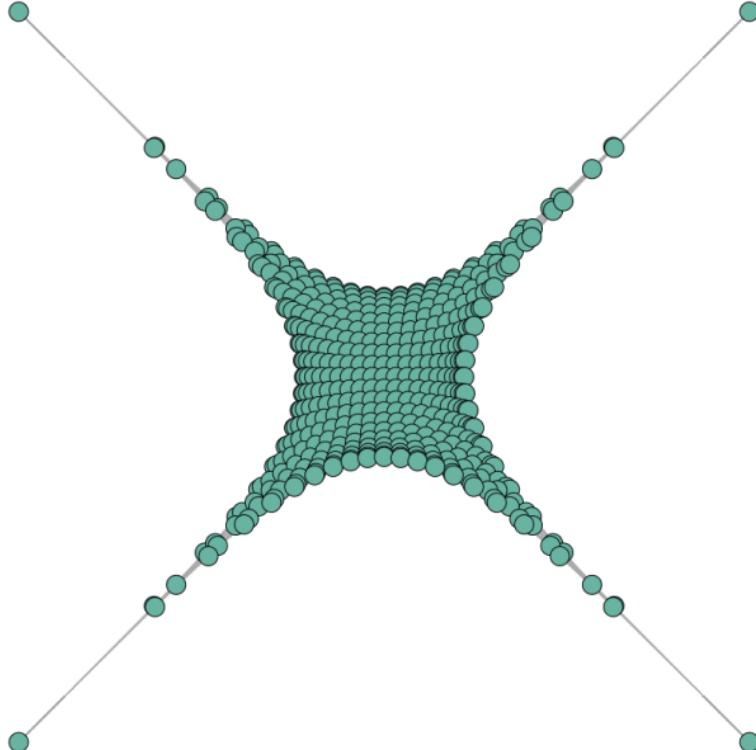


Figure 69: **L-BFGS**: 0: f without length & Constrained spring scenario: 2: Corners

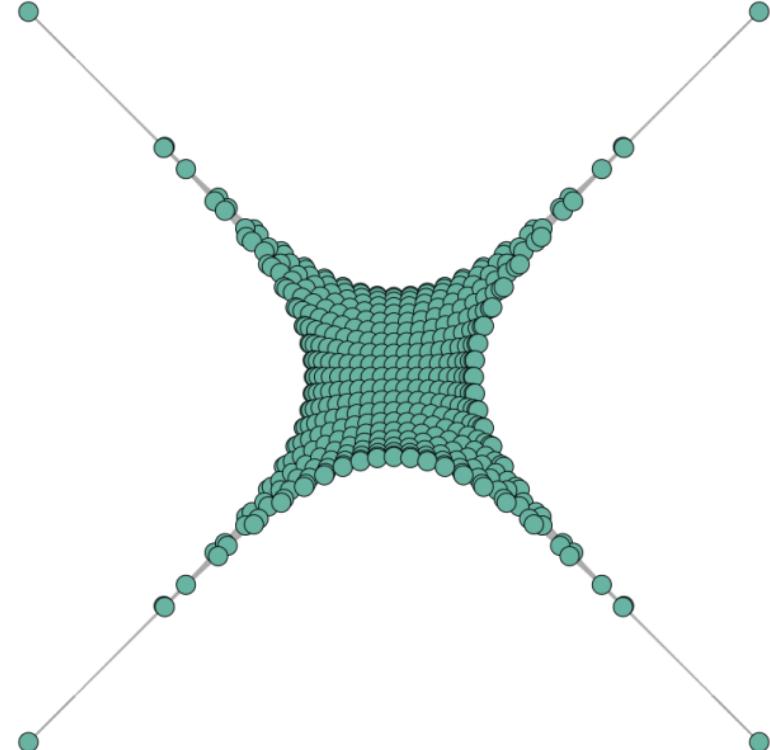


Figure 70: **Newton with projected Hessian**: 0: f without length & Constrained spring scenario: 2: Corners

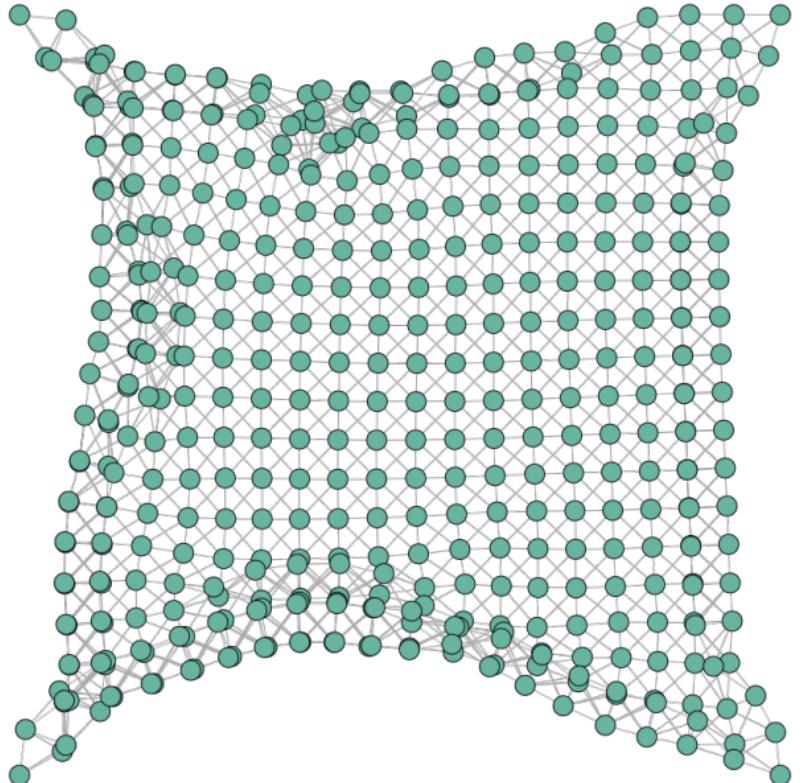


Figure 71: **Gauss Newton**: 1: f with length & Constrained spring scenario: 2: Corners

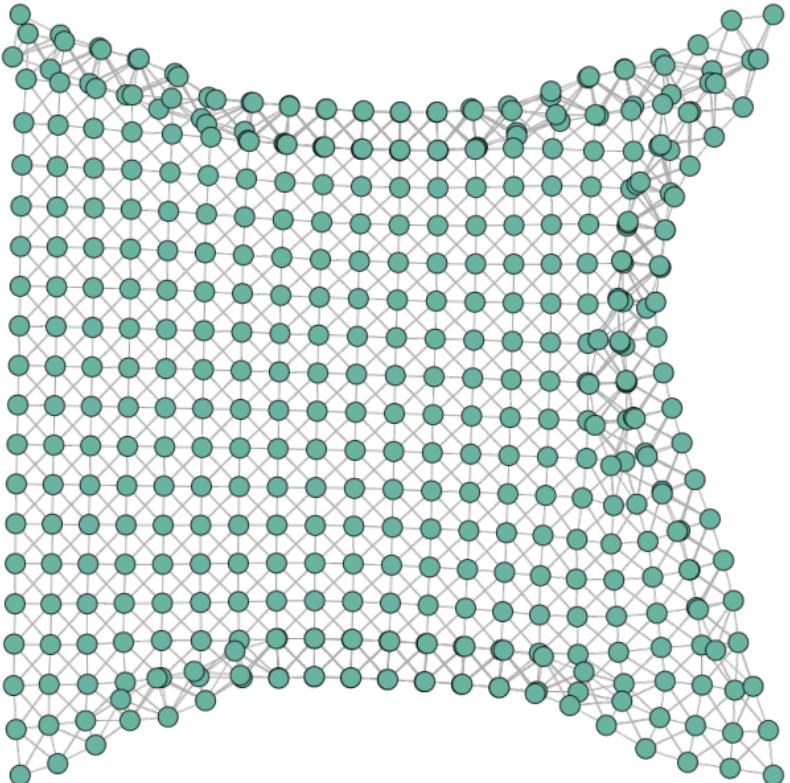


Figure 72: **Gradient Descent**: 1: f with length & Constrained spring scenario: 2: Corners

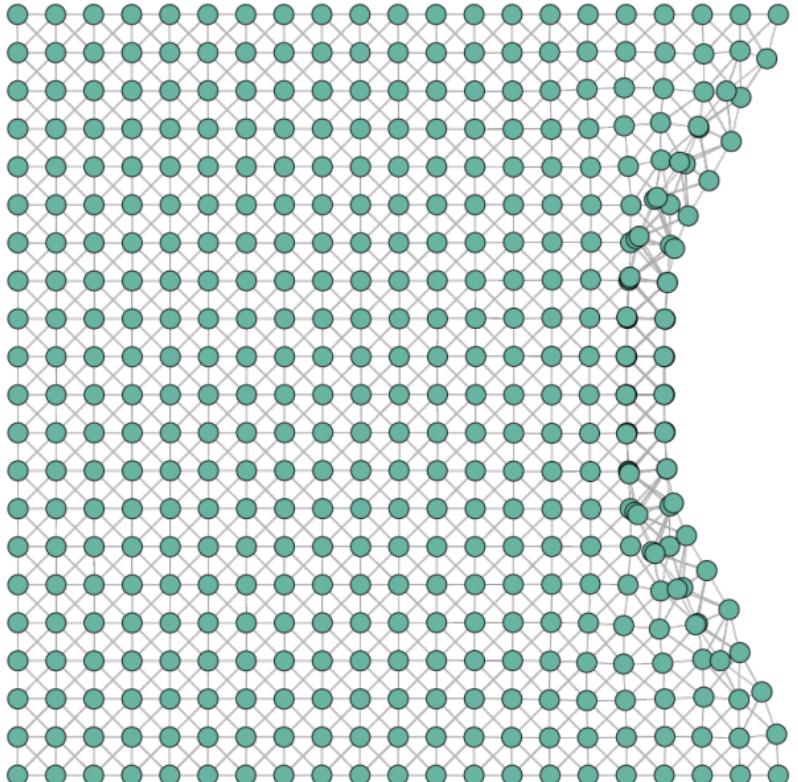


Figure 73: **Gradient Descent**: 1: f with length & Constrained spring scenario: 2: Corners

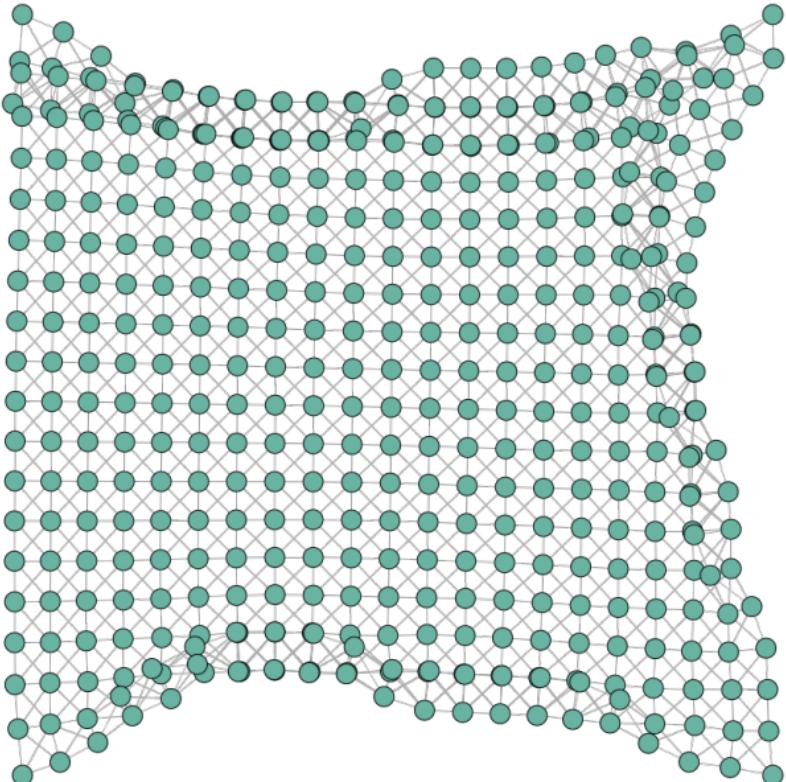


Figure 74: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

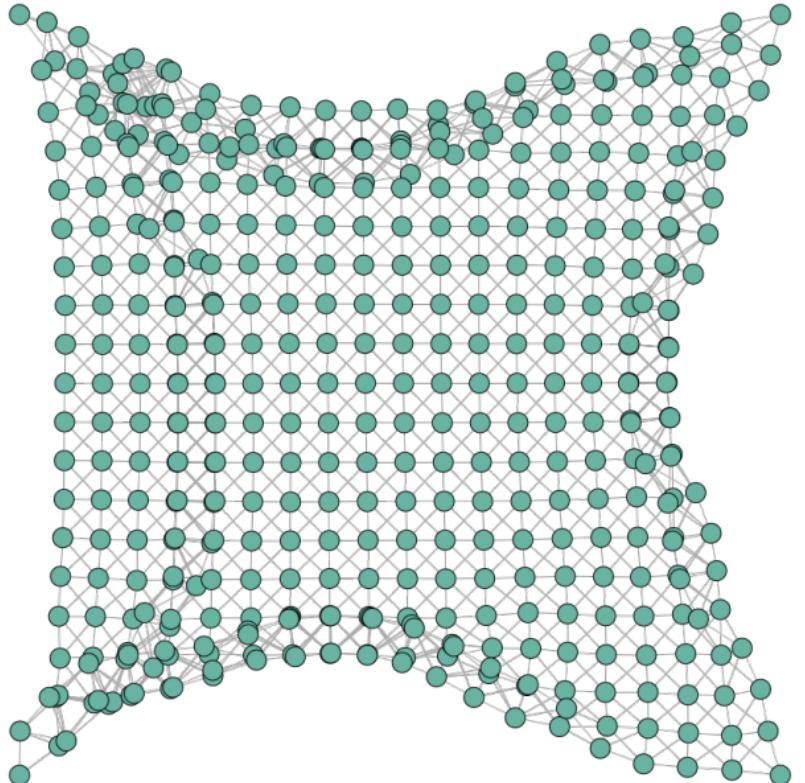


Figure 75: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

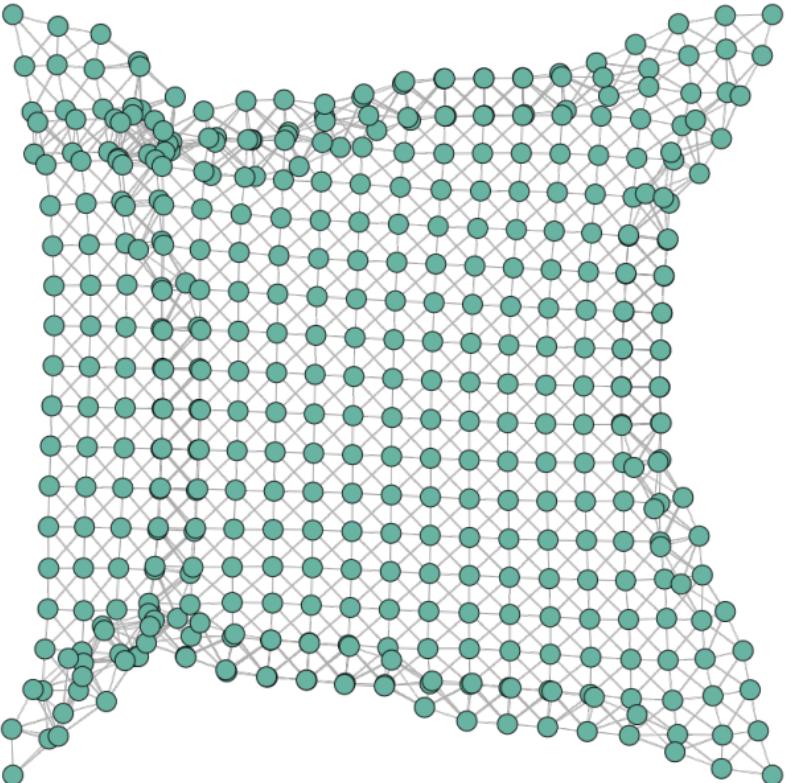


Figure 76: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

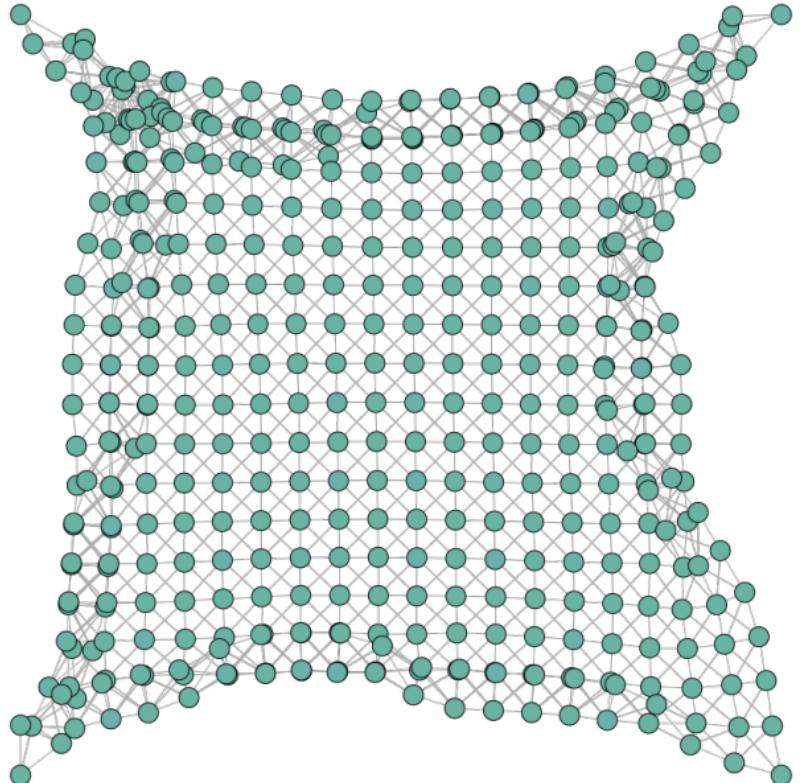


Figure 77: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

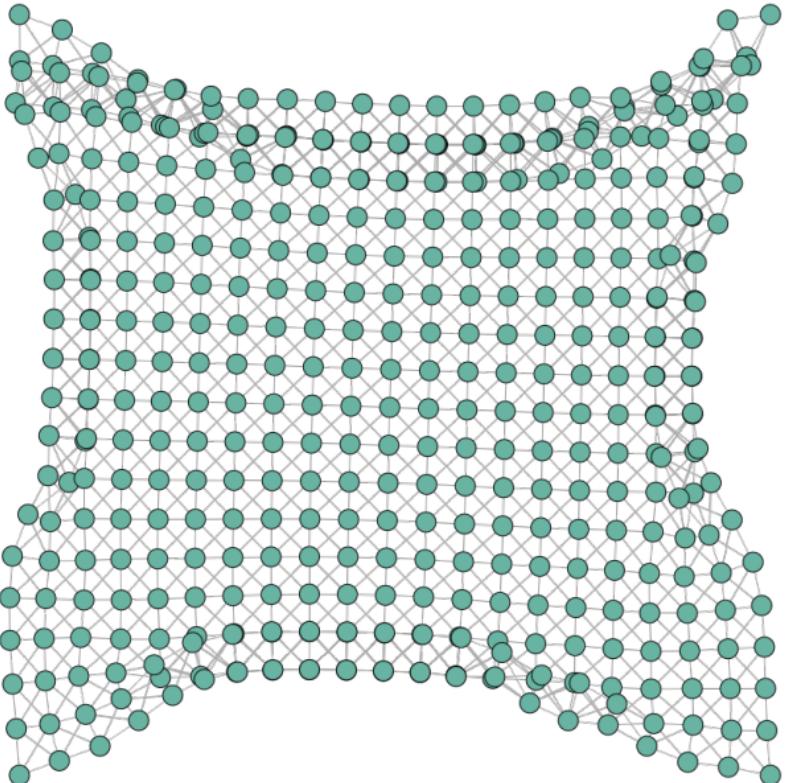


Figure 78: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

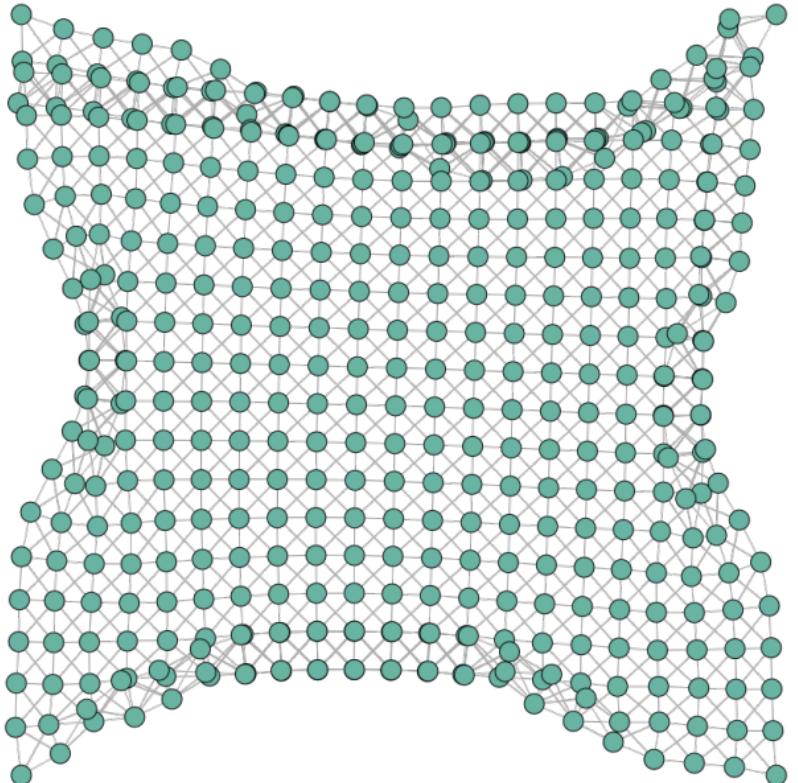


Figure 79: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

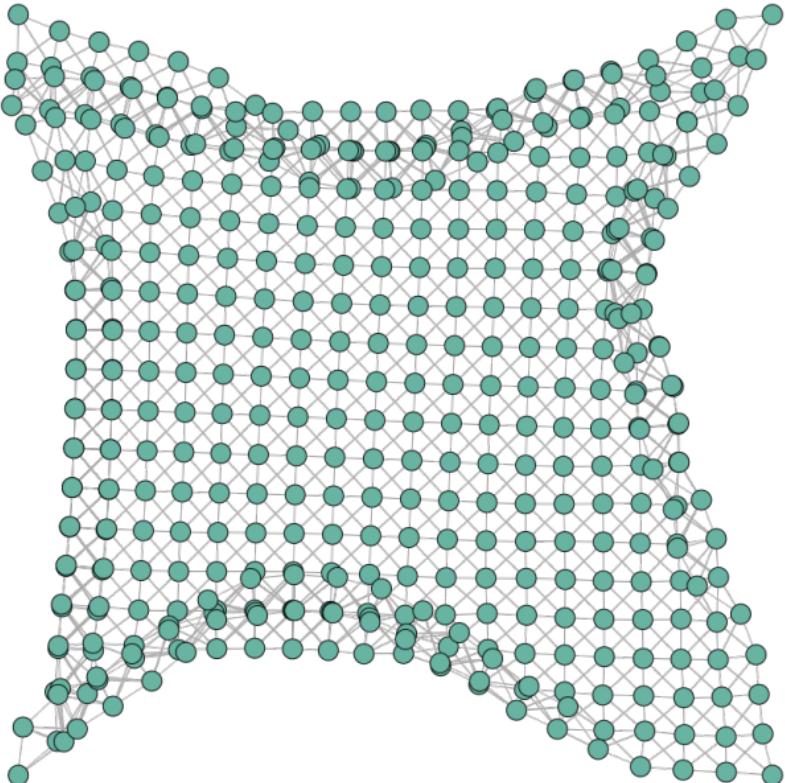


Figure 80: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

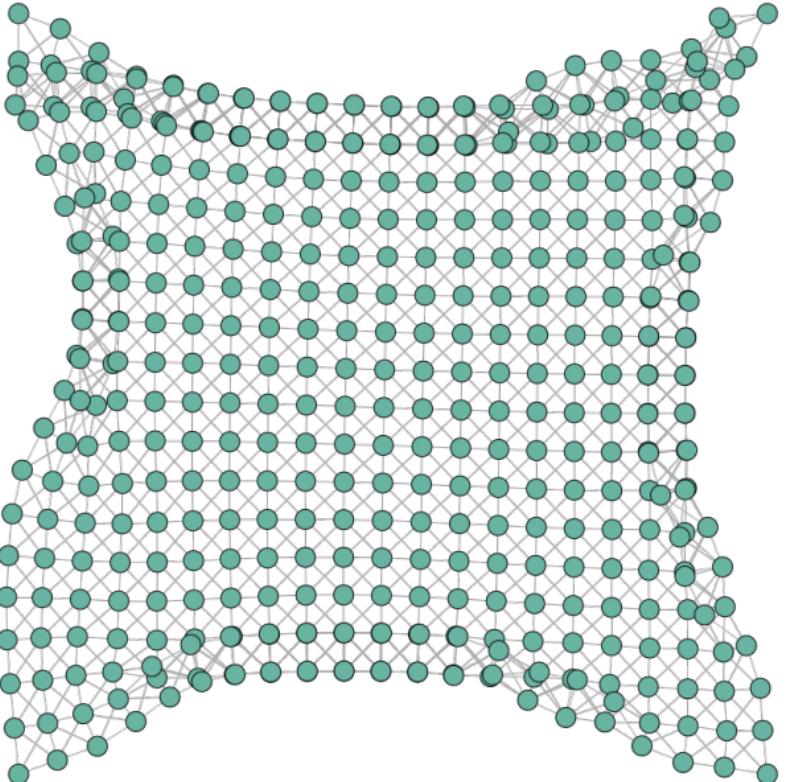


Figure 81: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

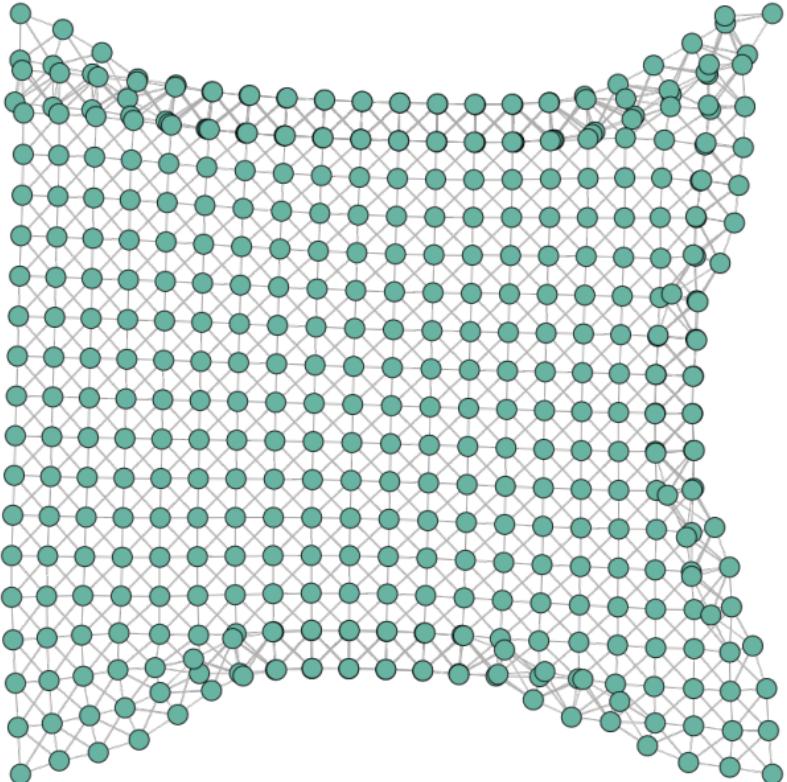


Figure 82: **L-BFGS**: 1: f with length & Constrained spring scenario: 2: Corners

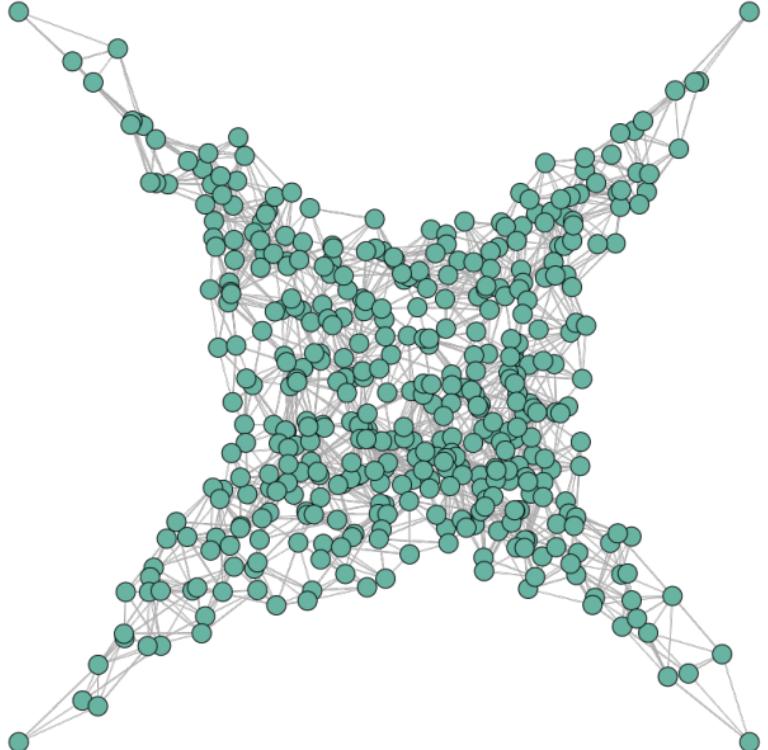


Figure 83: **Newton with projected Hessian:**
1: f with length & Constrained spring scenario:
2: Corners

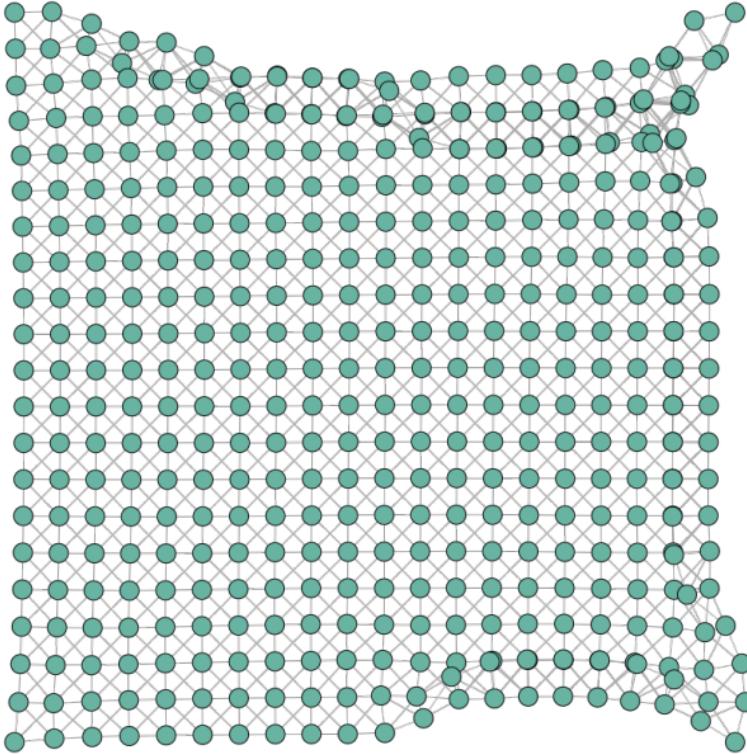


Figure 84: **Standard Newton:** 2: f with length
with positive local hessian & Constrained spring
scenario: 2: Corners

Executable	Arg1	Arg2	Function index description	Newton method	Spring constraint scenario	History (m)	Grid side	Iterations	Total time [s]	Total time evaluation [s]	Total time evaluation percentage [%]	eval_f time [s]	eval_f evaluations	eval_f avg [s]	eval_g time [s]	eval_g evaluations	eval_g avg [s]	eval_g factor	eval_h time [s]	eval_h evaluations	eval_h avg [s]	eval_h factor	
GaussNewton	0	-	0: f without length	Standard Newton	Corners	-	5	1	0.000747	0.000119	15.9304	0.00000	0	nan	0.00003	1	0.00003	nan	0.00009	1	0.00009	nan	
NewtonMethods	0	0	0: f without length	Standard Newton	Corners	-	5	2	0.000399	0.00017	42.6065	0.00001	4	0.00000	0.00000	2	0.00000	1.33333	0.00016	2	0.00008	53.33333	
GradientDescent	0	1	0: f without length	Standard Newton	Sides	-	5	12694	0.710848	0.509329	71.6509	0.49530	452808	0.00000	0.01403	12694	0.00000	1.01013	0.00000	0	nan	nan	
GradientDescent	0	2	0: f without length	Standard Newton	Sides	-	5	4481	0.270623	0.180096	66.5487	0.17502	159083	0.00000	0.00507	4481	0.00000	1.02901	0.00000	0	nan	nan	
LBFGS	0	8	0: f without length	Standard Newton	Corners	8	5	799	0.010789	0.004768	44.1932	0.00376	3054	0.00000	0.0101	800	0.00000	1.02084	0.00000	0	nan	nan	
LBFGS	0	9	0: f without length	Standard Newton	Corners	9	5	1228	0.015922	0.005962	37.445	0.00471	4610	0.00000	0.00126	1229	0.00000	1.00112	0.00000	0	nan	nan	
LBFGS	0	10	0: f without length	Standard Newton	Corners	10	5	859	0.011162	0.004125	36.9557	0.00326	3220	0.00000	0.00086	860	0.00000	0.98767	0.00000	0	nan	nan	
LBFGS	0	11	0: f without length	Standard Newton	Corners	11	5	1013	0.012681	0.004915	38.7588	0.00381	3754	0.00000	0.00110	1014	0.00000	1.06872	0.00000	0	nan	nan	
LBFGS	0	12	0: f without length	Standard Newton	Corners	12	5	719	0.008649	0.003729	43.1148	0.00289	2597	0.00000	0.00084	720	0.00000	1.04231	0.00000	0	nan	nan	
LBFGS	0	13	0: f without length	Standard Newton	Corners	13	5	952	0.012662	0.004656	36.7714	0.00368	3522	0.00000	0.00098	953	0.00000	0.98016	0.00000	0	nan	nan	
LBFGS	0	14	0: f without length	Standard Newton	Corners	14	5	852	0.011712	0.004078	34.819	0.00316	3113	0.00000	0.00092	853	0.00000	1.05722	0.00000	0	nan	nan	
LBFGS	0	15	0: f without length	Standard Newton	Corners	15	5	974	0.013623	0.004793	35.1831	0.00372	3615	0.00000	0.00107	975	0.00000	1.06432	0.00000	0	nan	nan	
LBFGS	0	16	0: f without length	Standard Newton	Corners	16	5	885	0.012294	0.004205	34.2037	0.00330	3193	0.00000	0.00091	886	0.00000	0.98972	0.00000	0	nan	nan	
NewtonMethods	2	0	0: f without length	Newton with projected Hessian	Corners	-	5	2	0.000262	0.0001	38.1679	0.00000	4	0.00000	0.00000	2	0.00000	1.50000	0.00009	2	0.00005	46.50000	
GaussNewton	1	-	1: f with length		Standard Newton	Corners	-	5	64	0.00656	0.002652	40.4268	0.00042	206	0.00000	0.00069	64	0.00001	5.31827	0.00155	64	0.00002	11.93330
GradientDescent	1	1	1: f with length		Corners	-	5	6269	0.39135	0.247737	63.3032	0.24112	221556	0.00000	0.00662	6269	0.00000	0.96987	0.00000	0	nan	nan	
GradientDescent	1	2	1: f with length		Corners	-	5	2157	0.143302	0.089039	62.1338	0.08669	76999	0.00000	0.00235	2157	0.00000	0.96600	0.00000	0	nan	nan	
LBFGS	1	8	1: f with length		Corners	8	5	1651	0.023795	0.010094	42.4207	0.00804	6153	0.00000	0.00205	1652	0.00000	0.94920	0.00000	0	nan	nan	
LBFGS	1	9	1: f with length		Corners	9	5	1875	0.022669	0.009164	40.4253	0.00726	7092	0.00000	0.00190	1876	0.00000	0.99144	0.00000	0	nan	nan	
LBFGS	1	10	1: f with length		Corners	10	5	1652	0.021032	0.007913	37.6236	0.00620	6124	0.00000	0.00172	1653	0.00000	1.02665	0.00000	0	nan	nan	
LBFGS	1	11	1: f with length		Corners	11	5	1791	0.023258	0.009073	39.0102	0.00710	6519	0.00000	0.00198	1792	0.00000	1.01287	0.00000	0	nan	nan	
LBFGS	1	12	1: f with length		Corners	12	5	1836	0.023967	0.009197	38.3736	0.00729	6665	0.00000	0.00191	1837	0.00000	0.94973	0.00000	0	nan	nan	
LBFGS	1	13	1: f with length		Corners	13	5	1839	0.023386	0.009075	38.8053	0.00709	6626	0.00000	0.00198	1840	0.00000	1.00625	0.00000	0	nan	nan	
LBFGS	1	14	1: f with length		Corners	14	5	1963	0.025993	0.009633	37.06	0.00751	7079	0.00000	0.00213	1964	0.00000	1.02077	0.00000	0	nan	nan	
LBFGS	1	15	1: f with length		Corners	15	5	1805	0.026782	0.010276	38.3691	0.00801	6493	0.00000	0.00227	1806	0.00000	1.01823	0.00000	0	nan	nan	
LBFGS	1	16	1: f with length		Corners	16	5	1848	0.02579	0.008677	33.6448	0.00683	6755	0.00000	0.00185	1849	0.00000	0.99067	0.00000	0	nan	nan	
NewtonMethods	2	1	1: f with length	Newton with projected Hessian	Standard Newton	Corners	-	5	8	0.000695	0.000219	31.5108	0.00002	21	0.00000	0.00001	8	0.00000	1.02717	0.00019	8	0.00002	21.34239
NewtonMethods	0	2	2: f with length with positive local hessian		Standard Newton	Corners	-	5	61	0.019804	0.015783	79.696	0.00000	186	0.00000	0.00006	61	0.00000	0.97382	0.01553	61	0.00025	247.94146
GaussNewton	0	-	0: f without length	Standard Newton	Corners	-	10	1	0.001136	0.000193	16.9894	0.00000	0	nan	0.00007	1	0.00007	nan	0.00012	1	0.00012	nan	
NewtonMethods	0	0	0: f without length	Standard Newton	Corners	-	10	3	0.001221	0.000397	32.5143	0.00004	8	0.00001	0.00002	3	0.00001	1.13333	0.00034	3	0.00011	22.66667	
GradientDescent	0	1	0: f without length	Standard Newton	Sides	-	10	48455	8.41591	7.73977	91.9659	7.52374	1729685	0.00000	0.21603	48455	0.00000	1.02497	0.00000	0	nan	nan	
GradientDescent	0	2	0: f without length	Standard Newton	Sides	-	10	15012	3.03963	2.66133	87.5546	2.57982	536060	0.00000	0.08151	15012	0.00001	1.12825	0.00000	0	nan	nan	
LBFGS	0	8	0: f without length	Standard Newton	Corners	8	10	2543	0.15415	0.057438	37.2611	0.04503	9470	0.00000	0.01241	2544	0.00000	1.02615	0.00000	0	nan	nan	