Data Visualization and Optimization

Minerva University

Supervisor

Prof. JLevitt

October 28$^{\text{th}}$, 2025

1 Problem 1: Area Model and Visualization

1.1 Model Variables and Constraints

The channel geometry is defined by three main lengths: the right wall a, the base b, and the total sheet width $W = 3\,\text{m}$. When the right wall is bent at an angle θ, its vertical height is $h = a\sin\theta$ and its horizontal projection is $a\cos\theta$.

Because the sheet cannot stretch, the total material must satisfy

$$W = b + a + h \quad \Rightarrow \quad b = W - a(1 + \sin\theta),$$

which links the base b to the chosen a and θ.

This condition ensures that any increase in a or θ reduces b, showing a direct trade-off between the wall size and the base. Physically feasible channels must satisfy $b \geq 0$, which implies

$$a(1 + \sin\theta) \leq W, \qquad 0 < \theta < 90^\circ.$$

These constraints define the range of valid geometric designs for the model.

1.2 Area Expression

The channel cross-section forms a trapezoid. The bottom width is b, the top width is $b + a\cos\theta$, and the height is $h = a\sin\theta$. The area of a trapezoid is half its height times the sum of its parallel sides:

$$A = \frac{1}{2}h(2b + a\cos\theta).$$

Substituting $h = a\sin\theta$ and $b = W - a(1 + \sin\theta)$ gives

$$A(a, \theta) = \frac{1}{2}a\sin\theta\big[\,2W - 2a(1 + \sin\theta) + a\cos\theta\,\big].$$

This expression relates the area directly to the design variables a and θ under the width constraint.

1.3 Visualization Method

The function $A(a, \theta)$ was calculated on a grid for $a \in (0, W)$ and $\theta \in (0, 90^\circ)$. Points that give $b = W - a(1 + \sin\theta) < 0$ were removed since they are not physically possible. Two plots were created to observe how $A(a, \theta)$ changes with different values of a and θ:

- A contour plot with curved lines connecting points of equal area and a dashed curve showing where $b = 0$.
- A surface plot that shows the same function as a 3D shape, helping to see the highest point clearly.

Both plots have labeled axes, use meters and degrees as units, and apply a clear color scale for comparison.

1.4 Interpretation of Results

In the contour plot, the area increases toward the center of the curved lines, where they close around a single point. This region is near $a \approx 1.0\,\text{m}$ and $\theta \approx 60^\circ$, showing where the channel holds the most water. The surface plot shows the same pattern from a different view: the highest point on the surface appears at about the same values of a and θ. Since there are no other high points or separate peaks, the maximum is unique within the range of possible designs.

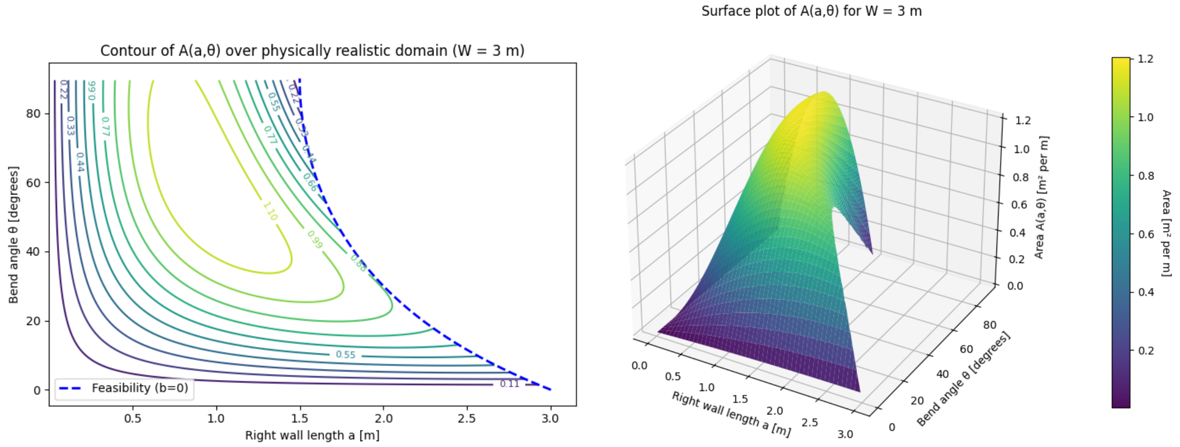

Figure 1: Left: Contour plot showing how the area changes with a and θ, with a dashed line marking the boundary $b = 0$. Right: Surface plot showing the same function as a 3D surface, where the highest point represents the largest cross-sectional area. Both figures show that the maximum occurs inside the feasible region, not along the boundary.

Problem 2: Analytical Optimization of $A(a, \theta)$

Recalling the area function from Part 1,

$$A(a, \theta) = \tfrac{1}{2} a \sin\theta \big[2W - 2a(1 + \sin\theta) + a\cos\theta\big],$$

with $W = 3\,\text{m}$, the goal is to find the exact values of a and θ that maximize the cross-sectional area.

To do this, I took the partial derivatives of A with respect to both a and θ and set them equal to zero:

$$\frac{\partial A}{\partial a} = (W - 2a\sin\theta + a\cos\theta - 2a)\sin\theta = 0, \qquad \frac{\partial A}{\partial \theta} = \tfrac{a}{2}\big(2W\cos\theta - 2a\sin 2\theta - 2a\cos\theta + a\cos 2\theta\big) = 0.$$

Substituting the first equation into the second yields

$$\cos\theta = \tfrac{1}{2} \quad \Rightarrow \quad \theta^* = \tfrac{\pi}{3}.$$

Solving these first-order conditions gives the exact symbolic solutions at $W = 3$:

$$a^* = \tfrac{2W}{2\sqrt{3}+3} = 4\sqrt{3} - 6, \qquad b^* = 3 - \sqrt{3}, \qquad h^* = 6 - 3\sqrt{3}, \qquad A^* = \frac{9}{2(2 + \sqrt{3})}.$$

and numerically,

$$a^* \approx 0.93\,\text{m}, \qquad \theta^* \approx 1.05\,\text{rad}\ (60^\circ).$$

These satisfy the physical condition $a(1+\sin\theta) \le W$, ensuring that the base $b = W - a(1+\sin\theta)$ remains positive and the channel geometry is valid.

To confirm that this stationary point represents a maximum, I evaluated the Hessian matrix:

$$H = \begin{bmatrix} A_{aa} & A_{a\theta} \\ A_{\theta a} & A_{\theta\theta} \end{bmatrix} = \begin{bmatrix} -2.80 & -1.50 \\ -1.50 & -1.55 \end{bmatrix}_{(a^*, \theta^*)}.$$

The determinant is $D = 2.09 > 0$, and $A_{aa} = -2.80 < 0$, so the matrix is negative definite, confirming a **local maximum**.

A short boundary check verifies that this point is also the **global maximum**:

- For $b = 0 \Rightarrow a = \frac{W}{1+\sin\theta}$, the area is

$$A(\theta) = \frac{W^2}{2} \frac{\sin\theta\cos\theta}{(1+\sin\theta)^2},$$

maximized at $\theta = 30^\circ$ with $A = 0.866 < A^*$.
- For $\theta = \frac{\pi}{2}$, the maximum is $A = W^2/8 = 1.125 < A^*$.
- As $a \to 0$ or $\theta \to 0$, $A \to 0$.

Hence the interior solution gives the largest possible area within the feasible region.

The maximum area is

$$A^* \approx 1.21\,\mathrm{m}^2,$$

and all derivatives were computed with θ in **radians**, while plots label θ in **degrees**.

Problem 3: Gradient Descent Optimization

To check that the analytic solution is correct, I used a gradient descent method with an **Armijo backtracking line search** on the area function $A(a, \theta)$. Since gradient descent normally minimizes, I used the negative form $f(a, \theta) = -A(a, \theta)$ so that minimizing f gives the same result as maximizing A. The algorithm also kept the condition $b = W - a(1 + \sin\theta) \geq 0$ at each step to make sure the updates stayed physically realistic. At each iteration, the descent direction was

$$d_k = -\frac{\nabla f(a_k, \theta_k)}{\|\nabla f(a_k, \theta_k)\|},$$

and the step length α_k satisfied the Armijo condition

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c\,\alpha_k \nabla f(x_k)^\top d_k,$$

with parameters $p = 0.5$, $c = 10^{-4}$, and $\alpha_0 = 1.0$. Termination occurred when $\|\nabla f\| < 10^{-6}$ or 500 iterations were reached.
The algorithm was initialized from three feasible points: $(1.5, 0.0)$, $(0.6, 0.4)$, and $(1.6, 1.0)$. All runs converged to the same solution:

$$a^* \approx 0.93 \text{ m}, \qquad \theta^* \approx 1.05 \text{ rad } (60^\circ), \qquad A^* \approx 1.206 \text{ m}^2/\text{m}.$$

Convergence took **21, 33, and 32 steps** respectively. All three paths converged smoothly to the same optimum, as shown in **Figure 2**.

Problem 3 (Part 2): Fixed-α GD and Two Additional Methods

Common setup All methods optimize the same objective $f(a, \theta) = -A(a, \theta)$ under the feasibility condition $b = W - a(1 + \sin\theta) \geq 0$ (projection applied if needed).
- **Starting points:** $(1.5, 0^\circ)$, $(0.6, 40^\circ)$, $(1.0, 75^\circ)$
- **Termination criterion:** $\|\nabla f\| < 10^{-6}$ or a maximum of 2000 iterations

(i) Gradient Descent with Fixed Learning Rate.
- **Learning rate:** $\alpha = 0.02$
- **Behavior:** Constant step size; sensitive to curvature and starting point
- **Result:** Zig-zag motion near $b = 0$ but converges to (a^*, θ^*)

(ii) Nonlinear Conjugate Gradient (PR+) with Armijo.
- **Direction update:** $d_k = -g_k + \beta_k d_{k-1}$
- **Line search:** Armijo with $c = 10^{-4}$, $\rho = 0.5$
- **Behavior:** Smoother progress, less boundary oscillation

(iii) Momentum Gradient Descent.

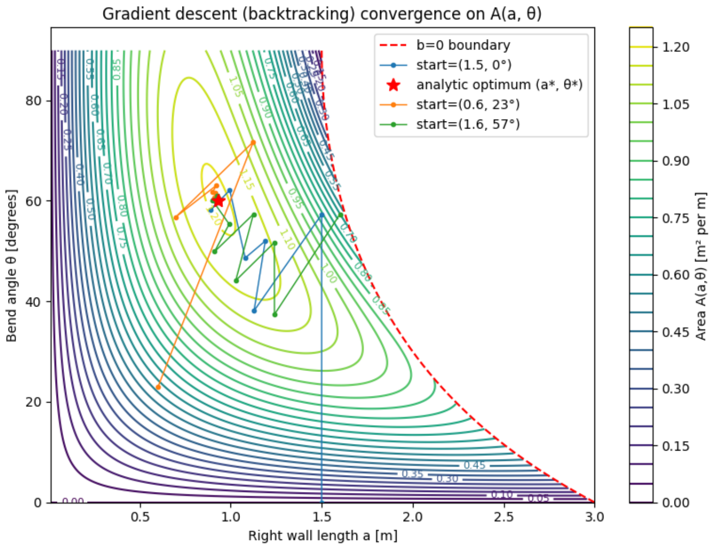

Figure 2: Gradient descent trajectories from three starting points. All paths stay inside the feasible region (left of the red dashed $b = 0$ line) and end at the same optimum (a^*, θ^*), confirming both the correctness of the analytic solution and the physical realism of the steps.

- **Update:** $v_{k+1} = \beta v_k - \alpha \nabla f(x_k),\ x_{k+1} = x_k + v_{k+1}$
- **Parameters:** $\alpha = 0.04,\ \beta = 0.9$
- **Behavior:** Faster early progress, mild overshoot near $b = 0$

Trajectory Plots. To conserve space, the first two plots are shown side by side, followed by the third below.

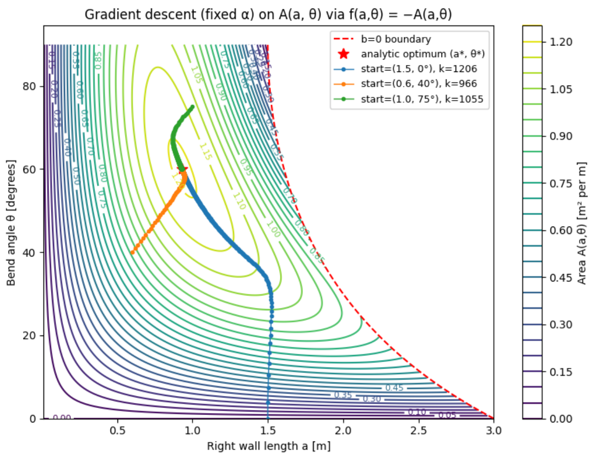

Figure 3: Fixed-α GD ($\alpha = 0.02$). Trajectories from three starts; dashed curve shows $b = 0$; star marks the analytic maximizer.

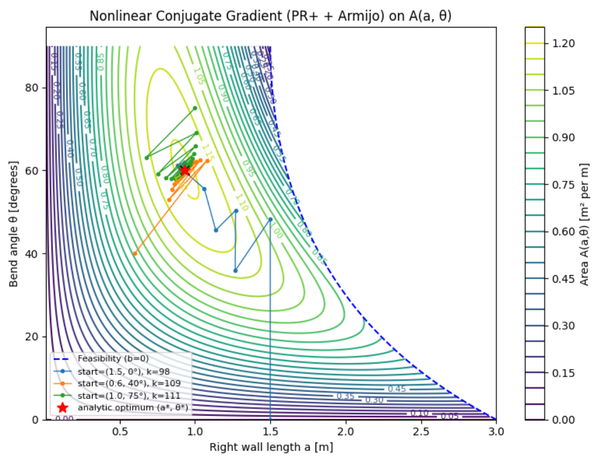

Figure 4: NLCG (PR+ Armijo) trajectories from same starts showing smoother convergence.

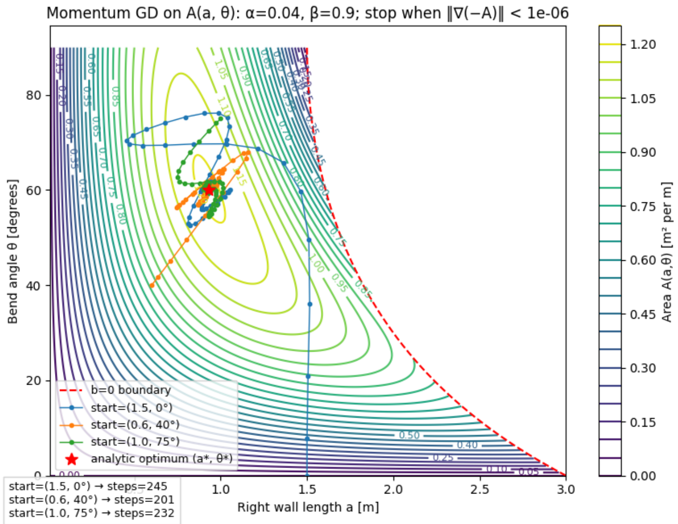

Figure 5: Momentum GD ($\alpha = 0.04,\ \beta = 0.9$) trajectories on $A(a, \theta)$ from the same starts.

Problem 3 (Part 3): Performance Summary

Metric. All results are reported in terms of *outer iterations* ("steps"). For methods using backtracking, each outer step may include several internal function or gradient evaluations.

Table 1: Number of outer iterations required to reach $\|\nabla f\| < 10^{-6}$ from each starting point. All methods converge to $A^* \approx 1.20577$ m^2/m.

Method	**(1.5, 0°)**	**(0.6, 40°)**	**(1.0, 75°)**	**Mean**
GD + Armijo (Part 3 main)	21	33	32	28.7
Fixed-α GD ($\alpha = 0.02$)	1206	966	1055	1075.7
NLCG (PR$^+$ + Armijo)	98	109	111	106.0
Momentum GD ($\alpha = 0.04,\ \beta = 0.9$)	245	201	232	226.0

Summary. Gradient Descent with Armijo backtracking required the fewest *outer iterations*, as the line search adaptively selected step sizes suited to local curvature. The Nonlinear Conjugate Gradient (PR$^+$) method produced smooth, stable progress with moderate step counts. Momentum GD accelerated early progress but required parameter tuning to prevent oscillations near the $b = 0$ boundary. Fixed-α GD was the most sensitive to step size and needed over 10^3 steps to converge. All methods reached the same interior maximizer.

Obstacles Encountered.

- Near the feasibility boundary, I observed oscillations in both Fixed-α GD and Momentum GD; applying projection and reducing α helped stabilize the updates.
- In NLCG, I had to restart frequently when $d_k^\top g_k \geq 0$; using the PR$^+$ rule improved direction stability and convergence speed.
- Comparing methods was difficult because "outer iterations" do not reflect the extra backtracking evaluations in line-search methods.

AI Assistance Acknowledgment

I used ChatGPT ("GPT thinking mode") to help with generating the plots in this report. I provided my own working code from CS164 Sessions 5–9 and asked GPT to ensure that the plotting matched the assignment instructions and visual formatting requirements. I also used GPT for assistance with organizing and formatting the LaTeX document. All analytical derivations, parameter selections, and interpretations were done by me. I reviewed and tested the code output myself, added comments, and confirmed that each plot was consistent with the analytical results. Copilot was also helpful, as it made code completion easier through analytical calculations. I added a docstring so that it's easier to track for each code cell, and this is what we have used for the CS110 class before. I wanted to work on Newton's method, but I still think I don't fully understand it, and that's why I went with the other two methods. I also used Grammarly and QuillBot to check grammar, spelling, and punctuation mistakes.

Note: The complete Python implementation used to generate the figures is submitted along with this report as a secondary uploaded file (`CS164_Assignment2_Notebook.ipynb`). A public version of the same notebook is also hosted on Google Colab: `CS164 Assignment 2`.

Appendix

Part 2: Lagrange Multiplier Method

This section includes the equality-constrained optimization approach used to confirm the results obtained in Problem 2. The area function was defined as

$$A(a, b, \theta) = \tfrac{1}{2} a \sin\theta (2b + a\cos\theta),$$

subject to the sheet-width constraint

$$g(a, b, \theta) = a + b + a\sin\theta - W = 0,$$

where $W = 3\,\mathrm{m}$.

A Lagrangian function $L = A - \lambda g$ was formed, and the partial derivatives with respect to a, b, θ, and λ were set to zero to find stationary points. Symbolically, several mathematical roots were obtained, but most were physically invalid (e.g., negative a, negative b, or $\theta = 0$). After applying the feasibility condition

$$a > 0, \quad b > 0, \quad 0 < \theta < \tfrac{\pi}{2},$$

only one valid stationary point remained:

$$a^* \approx 0.93\,\mathrm{m}, \quad b^* \approx 1.27\,\mathrm{m}, \quad \theta^* \approx 1.05\,\mathrm{rad}\,(60.1^\circ), \quad \lambda \approx 0.80.$$

This solution satisfies the constraint exactly, with $g(a^*, b^*, \theta^*) \approx 2.2 \times 10^{-16}$, confirming numerical consistency. It also meets the feasibility requirement $a(1 + \sin\theta) \leq W$, keeping b positive and ensuring a physically realistic geometry.

The corresponding maximum area is

$$A^* = 1.21\,\mathrm{m}^2/\mathrm{m},$$

which matches the value obtained in Problem 2 through direct differentiation. Both approaches therefore converge to the same global maximum within the feasible region, confirming the analytical consistency between the unconstrained and constrained methods.

Part 3: Gradient Descent Iteration Tables and Results

This section includes the numerical convergence details for the gradient descent implementation used in Problem 3. The results summarize the evolution of a, θ, and the corresponding area $A(a, \theta)$ during the optimization process. All runs enforced the feasibility condition $b = W - a(1 + \sin\theta) \geq 0$, ensuring the geometry stayed physically realistic throughout.

Each run began from a different feasible starting point and used Armijo backtracking with parameters $p = 0.5$, $c = 10^{-4}$, and $\alpha_0 = 1.0$. Convergence was achieved when $\|\nabla f\| < 10^{-6}$ or a maximum of 500 iterations was reached. The first ten iterations of each run are shown below.

Iter	a **[m]**	θ **[rad]**	$A(a,\theta)$ [m^2/m]	$\|\nabla f\|$	α
0	1.5000	0.0000	0.0000	3.38e+00	1.00e+00
1	1.5000	1.0000	0.8116	1.94e+00	5.00e-01
2	1.1284	0.6655	1.1278	5.94e-01	2.50e-01
3	1.1901	0.9077	1.1614	4.68e-01	1.25e-01
4	1.0798	0.8490	1.1885	1.48e-01	2.50e-01
5	0.9951	1.0842	1.1945	3.00e-01	1.25e-01
6	0.8905	1.0157	1.2013	1.81e-01	6.25e-02
7	0.9418	1.0515	1.2054	5.22e-02	1.56e-02
8	0.9284	1.0433	1.2058	7.74e-03	3.91e-03
9	0.9311	1.0462	1.2058	7.02e-03	1.95e-03

Table 2: First 10 iterations for start $(1.5, 0.0)$. Converged in 21 steps.

Iter	a **[m]**	θ **[rad]**	$A(a,\theta)$ [m^2/m]	$\|\nabla f\|$	α
0	0.6000	0.4000	0.5707	1.40e+00	1.00e+00
1	1.1240	1.2517	1.0513	1.15e+00	5.00e-01
2	0.6975	0.9907	1.1144	7.65e-01	2.50e-01
3	0.9219	1.1009	1.2041	9.36e-02	3.13e-02
4	0.9009	1.0777	1.2052	3.33e-02	1.56e-02
5	0.9163	1.0750	1.2055	2.55e-02	1.56e-02
6	0.9116	1.0602	1.2056	2.81e-02	1.56e-02
7	0.9270	1.0628	1.2056	2.98e-02	1.56e-02
8	0.9165	1.0513	1.2056	2.91e-02	1.56e-02
9	0.9309	1.0572	1.2056	2.99e-02	1.56e-02

Table 3: First 10 iterations for start $(0.6, 0.4)$. Converged in 33 steps.

Iter	a **[m]**	θ **[rad]**	$A(a,\theta)$ [m^2/m]	$\|\nabla f\|$	α
0	1.6000	1.0000	0.6542	2.37e+00	5.00e-01
1	1.2405	0.6524	1.1291	4.52e-01	2.50e-01
2	1.2399	0.9024	1.1396	6.08e-01	2.50e-01
3	1.0278	0.7704	1.1632	4.50e-01	2.50e-01
4	1.1247	1.0073	1.1664	5.29e-01	2.50e-01
5	0.9113	0.8706	1.1745	4.41e-01	1.25e-01
6	0.9917	0.9663	1.2029	5.71e-02	1.25e-01
7	0.8987	1.0498	1.2047	8.79e-02	3.13e-02
8	0.9267	1.0636	1.2056	3.08e-02	1.56e-02
9	0.9163	1.0521	1.2056	2.82e-02	7.81e-03

Table 4: First 10 iterations for start $(1.6, 1.0)$. Converged in 32 steps.

Across all runs, the area $A(a, \theta)$ consistently increased and stabilized around $A^* \approx 1.206$ m^2/m with $\|\nabla f\| \to 0$, confirming convergence to the same optimum found analytically. The small variation in iteration counts reflects the different initial gradients and proximity to the feasible boundary, but all trajectories reached the same physically meaningful maximum.

Part 3 (Reference): Methods and Conditions Used

Objective and Feasibility.

$$f(a,\theta) = -A(a,\theta), \qquad b = W - a(1 + \sin\theta) \geq 0.$$

Termination (all methods).

$$\|\nabla f(x_k)\| < 10^{-6} \quad \text{or} \quad k \text{ hits the iteration cap used in the run.}$$

(i) Fixed-α Gradient Descent (GD).

$$x_{k+1} = x_k - \alpha \nabla f(x_k), \qquad \alpha = 0.02.$$

(ii) Nonlinear Conjugate Gradient (PR$^+$) + Armijo.

$$\begin{aligned} &\text{Search dir:} \qquad d_k = -g_k + \beta_k d_{k-1}, \quad g_k := \nabla f(x_k),\ d_0 = -g_0, \\ &\text{PR}^+ \text{ coefficient:} \qquad \beta_k = \max\left(0,\ \frac{g_k^\top (g_k - g_{k-1})}{g_{k-1}^\top g_{k-1}}\right). \end{aligned}$$

$$\text{Restart if } d_k^\top g_k \geq 0 \ \Rightarrow\ d_k = -g_k.$$

$$\begin{aligned} &\text{Armijo backtracking: choose the largest } \alpha_k \in \{\alpha_0, \rho\alpha_0, \rho^2\alpha_0, \dots\} \\ &\text{s.t.} \quad f(x_k + \alpha_k d_k) \leq f(x_k) + c\,\alpha_k\, g_k^\top d_k, \quad c = 10^{-4},\ \rho = 0.5,\ \alpha_0 = 1.0. \end{aligned}$$

(iii) Momentum Gradient Descent (Heavy-Ball).

$$v_{k+1} = \beta v_k - \alpha \nabla f(x_k), \qquad x_{k+1} = x_k + v_{k+1},$$

$$\alpha = 0.04, \qquad \beta = 0.9, \qquad v_0 = 0.$$