1D Root Finding
Part II Roadmap

- Part I – Linear Algebra (units 1-12) \( Ac = b \)
- Part II – Optimization (units 13-20)
  - (units 13-16) Optimization -> Nonlinear Equations -> 1D roots/minima
  - (units 17-18) Computing/Avoiding Derivatives
  - (unit 19) Hack 1.0: “I give up” \( H = I \) and \( J \) is mostly 0 (descent methods)
  - (unit 20) Hack 2.0: “It’s an ODE!?” (adaptive learning rate and momentum)
Fixed Point Iteration

• Find roots of $g(t)$, where $g(t) = 0$

• Let $\hat{g}(t) = g(t) + t$ and iterate $t^{q+1} = \hat{g}(t^q)$ until convergence

• The converged $t^*$ satisfies $t^* = \hat{g}(t^*) = g(t^*) + t^*$, and so $g(t^*) = 0$

• Converges when $|g'(t^*)| < 1$ for a close enough initial guess (when $g$ is sufficiently smooth)

• $e^{q+1} = t^{q+1} - t^* = \hat{g}(t^q) - \hat{g}(t^*) = g'(\hat{t})(t^q - t^*) = g'(\hat{t})e^q$ for some $\hat{t}$ between $t^{q+1}$ and $t^*$ (by the Mean Value Theorem)

• When all $g'(\hat{t})$ have $|g'(\hat{t})| \leq C < 1$, then $|e^q| \leq C^q|e^0|$ proves convergence
Convergence Rate

• Consider \( \|e^{q+1}\| \leq C \|e^q\|^p \) as \( q \to \infty \) where \( C \geq 0 \)
  • When \( p = 1 \), \( C < 1 \) is required, and the convergence rate is \text{linear}.
  • When \( p > 1 \), the convergence rate is \text{superlinear}.
  • When \( p = 2 \), the convergence rate is \text{quadratic}.

• Statements only apply asymptotically (once convergence is happening).
• No guarantee of converging to a desired root (when other roots are present).
• Recall, \( g(t) = 0 \) may contain approximations, so it’s not clear how accurate the root finder needs to be anyways.
1D Newton’s Method

• Solve $g'(t^q)\Delta t = -g(t^q)$ and update $t^{q+1} = t^q + \Delta t = t^q - \frac{g(t^q)}{g'(t^q)}$

• Stop when $|g(t^q)| < \epsilon$, which implies $|t^{q+1} - t^q| < \frac{\epsilon}{|g'(t^q)|}$
  • Thus, poorly conditioned when $g'(t^*)$ is small
  • Especially problematic for repeated roots where $g'(t^*) = 0$

• Quadratic convergence rate ($p = 2$)

• Requires computing $g$ and $g'$ every iteration, and computing derivatives isn’t always straightforward/cheap (see units 17/18 Computing/Avoiding Derivatives)
1D Newton’s Method

- \( t^{q+1} = t^q - \frac{g(t^q)}{g'(t^q)} \) or alternatively \( g'(t^q) = \frac{\Delta g}{\Delta t} = \frac{g(t^q) - 0}{t^q - t^{q+1}} \)
Secant Method

- Replace $g'(t^q)$ in Newton’s method with an estimate (a few choices for this)
- Standard technique/method draws a line through previous iterates
- Estimate $g'(t^q) \approx \frac{g(t^q) - g(t^{q-1})}{t^q - t^{q-1}}$
- Then $t^{q+1} = t^q - g(t^q) \frac{t^q - t^{q-1}}{g(t^q) - g(t^{q-1})}$

- Superlinear convergence rate with $p \approx 1.618$
- Often/typically faster than Newton, since only $g$ (not $g'$) is needed while only a few extra iterations are required for the same accuracy
Secant Method

\[ t^{q+1} = t^q - g(t^q) \frac{t^q - t^{q-1}}{g(t^q) - g(t^{q-1})} \]

based on \( g'(t^q) \approx \frac{g(t^q) - g(t^{q-1})}{t^q - t^{q-1}} \)
Bisection Method

• If \( g(t_L)g(t_R) < 0 \), then (assuming continuity) the sign change indicates a root in the interval \([t_L, t_R]\)

• Let \( t_M = \frac{t_L + t_R}{2} \), and if \( g(t_L)g(t_M) < 0 \), set \( t_R = t_M \)
  • Otherwise, set \( t_L = t_M \) knowing that \( g(t_M)g(t_R) < 0 \) is true

• Iterate until \( t_R - t_L < \epsilon \)

• Guaranteed to converge to a root in the interval (unlike Newton/Secant)

• The interval shrinks in size by a factor of two each iteration

• So, linear convergence rate \( (p = 1) \) with \( C = \frac{1}{2} \)
Bisection Method

• If $g(t_L)g(t_M) < 0$, set $t_R = t_M$; otherwise, set $t_L = t_M$
Mixed Methods

• Given an interval with a root indicated by \( g(t_L)g(t_R) < 0 \)
• Iterate with Newton/Secant as long as the iterates stay inside the interval
  • When iteration attempts to leave the interval, use prior iterates to shrink the interval as much as possible (while still guaranteeing a root)
• Bisection can be used to continue to shrink the interval, whenever Newton/Secant would fail to stay inside the current interval
• Leverages the speed of Newton/Secant, while still guaranteeing convergence via Bisection
• Many/various strategies exist
Function/Derivative Requirements

• All methods require evaluation of the function $g$

• Newton also requires the derivative $g'$ (as do mixed methods using Newton)
Useful Derivatives

- \( \frac{\partial}{\partial t} c^{q+1}(t) = \Delta c^q \), since \( c^{q+1}(t) = c^q + t \Delta c^q \)

- \( \frac{\partial}{\partial t} F(c^{q+1}(t)) = J_F(c^{q+1}(t)) \Delta c^q \) and \( \frac{\partial}{\partial t} F^T(c^{q+1}(t)) = (\Delta c^q)^T J_F^T(c^{q+1}(t)) \)
  - \( \frac{\partial}{\partial t} F_i(c^{q+1}(t)) = (J_F)_i(c^{q+1}(t)) \Delta c^q \) where \( F_i(c^{q+1}(t)) \) are the scalar row components of \( F(c^{q+1}(t)) \)

- Scalar \( \hat{f}(c^{q+1}(t)) \) has system \( J_{\hat{f}}^T(c^{q+1}(t)) = 0 \) for critical points

- \( \frac{\partial}{\partial t} J_{\hat{f}}^T(c^{q+1}(t)) = H_{\hat{f}}^T(c^{q+1}(t)) \Delta c^q \) and \( \frac{\partial}{\partial t} J_{\hat{f}}(c^{q+1}(t)) = (\Delta c^q)^T H_{\hat{f}}(c^{q+1}(t)) \)
  - \( \frac{\partial}{\partial t} \left( J_{\hat{f}}^T \right)_i(c^{q+1}(t)) = (H_{\hat{f}}^T)_i(c^{q+1}(t)) \Delta c^q \)
Nonlinear Systems Problems

• Solve $J_F(c^q)\Delta c^q = -F(c^q)$ for $\Delta c^q$ and use $c^{q+1}(t) = c^q + t\Delta c^q$ in $F(c^{q+1}(t)) = 0$

• **Option 1**: For vector valued $F(c^{q+1}(t))$, find simultaneous (for all $i$) roots for all the $g_i(t) = F_i(c^{q+1}(t)) = 0$
  
  • Here, $g_i'(t) = (J_F)_i(c^{q+1}(t))\Delta c^q$

• **Option 2**: Find roots of $g(t) = \frac{1}{2}F^T(c^{q+1}(t))F(c^{q+1}(t)) = 0$
  
  • Here, $g'(t) = \frac{1}{2}F^T(c^{q+1}(t))J_F(c^{q+1}(t))\Delta c^q + \frac{1}{2}(\Delta c^q)^TJ_F^T(c^{q+1}(t))F(c^{q+1}(t))$
  
  • Both terms are scalars, so $g'(t) = F^T(c^{q+1}(t))J_F(c^{q+1}(t))\Delta c^q$
Optimization Problems

• Solve $H^T_{\hat{f}}(c^q)\Delta c^q = -J^T_{\hat{f}}(c^q)$ for $\Delta c^q$ and use $c^{q+1}(t) = c^q + t\Delta c^q$ in $J^T_{\hat{f}}(c^{q+1}(t)) = 0$

• **Option 1**: For vector valued $J^T_{\hat{f}}(c^{q+1}(t))$, find simultaneous (for all $i$) roots for all the $g_i(t) = (J^T_{\hat{f}})_i(c^{q+1}(t)) = 0$ to find the critical points of $\hat{f}(c)$
  
  • Here, $g_i'(t) = (H^T_{\hat{f}})_i(c^{q+1}(t))\Delta c^q$

• **Option 2**: Find roots of $g(t) = \frac{1}{2}J^T_{\hat{f}}(c^{q+1}(t))J^T_{\hat{f}}(c^{q+1}(t)) = 0$ to find or make progress toward critical points of $\hat{f}(c)$
  
  • Here, $g'(t) = \frac{1}{2}J^T_{\hat{f}}(c^{q+1}(t))H^T_{\hat{f}}(c^{q+1}(t))\Delta c^q + \frac{1}{2}(\Delta c^q)^TH^T_{\hat{f}}(c^{q+1}(t))J^T_{\hat{f}}(c^{q+1}(t))$
  • Both terms are scalars, so $g'(t) = J^T_{\hat{f}}(c^{q+1}(t))H^T_{\hat{f}}(c^{q+1}(t))\Delta c^q$

• **Option 3**: Minimize $\hat{f}(c^{q+1}(t))$ directly (see unit 16)