# 1D Root Finding

# Part II Roadmap

- Part I – Linear Algebra (units 1-12) $Ac = b$

linearize

line search

- Part II – Optimization (units 13-20)
  - (units 13-16) Optimization -> Nonlinear Equations -> 1D roots/minima ← Theory
  - (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)

Methods (units 17-20)

# Fixed Point Iteration

- Find roots of $g(t)$, i.e. where $g(t) = 0$

- Let $\hat{g}(t) = g(t) + t$ and iterate $t^{q+1} = \hat{g}(t^q)$ until convergence
- A converged $t^*$ satisfies $t^* = \hat{g}(t^*) = g(t^*) + t^*$ implying that $g(t^*) = 0$

- Converges when: $|g'(t^*)| < 1$, the initial guess is close enough to $t^*$, and $g$ 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 <u>linear</u>
  - When $p > 1$, the convergence rate is <u>superlinear</u>
  - When $p = 2$, the convergence rate is <u>quadratic</u>

- Statements only apply asymptotically (once convergence is happening)
- Might converge to a different non-desired root (when other roots are present)

- Solving $g(t) = 0$ may only approximate the problem being solved, so it's not clear how accurate the root finder needs to be anyways

# 1D Newton's Method

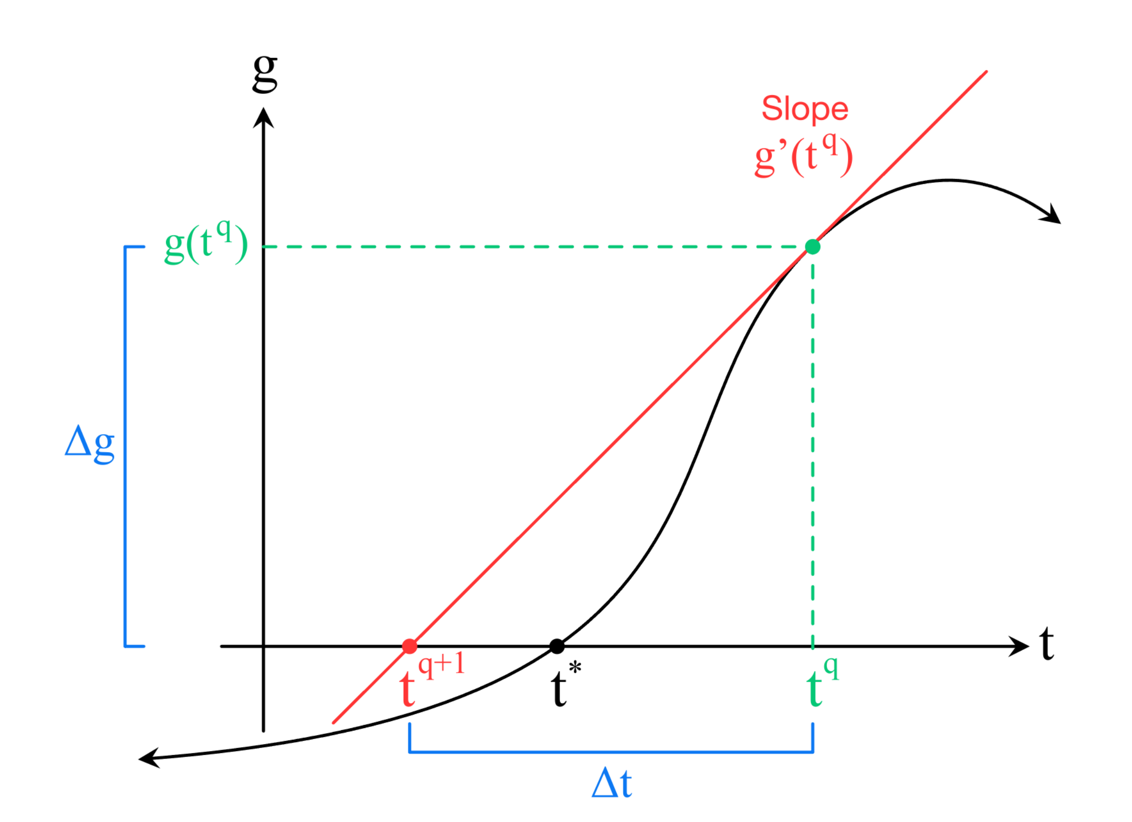
- Solve $g'(t^q)\Delta t = -g(t^q)$ and update $t^{q+1} = t^q + \Delta t = t^q - \frac{g(t^q)}{g'(t^q)}$

- 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 ($p = 2$), when not degenerate
- Requires computing $g$ and $g'$ every iteration; but, computing derivatives isn't always straightforward/cheap (see units 17/18 on Computing/Avoiding Derivatives)

# 1D Newton's Method

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



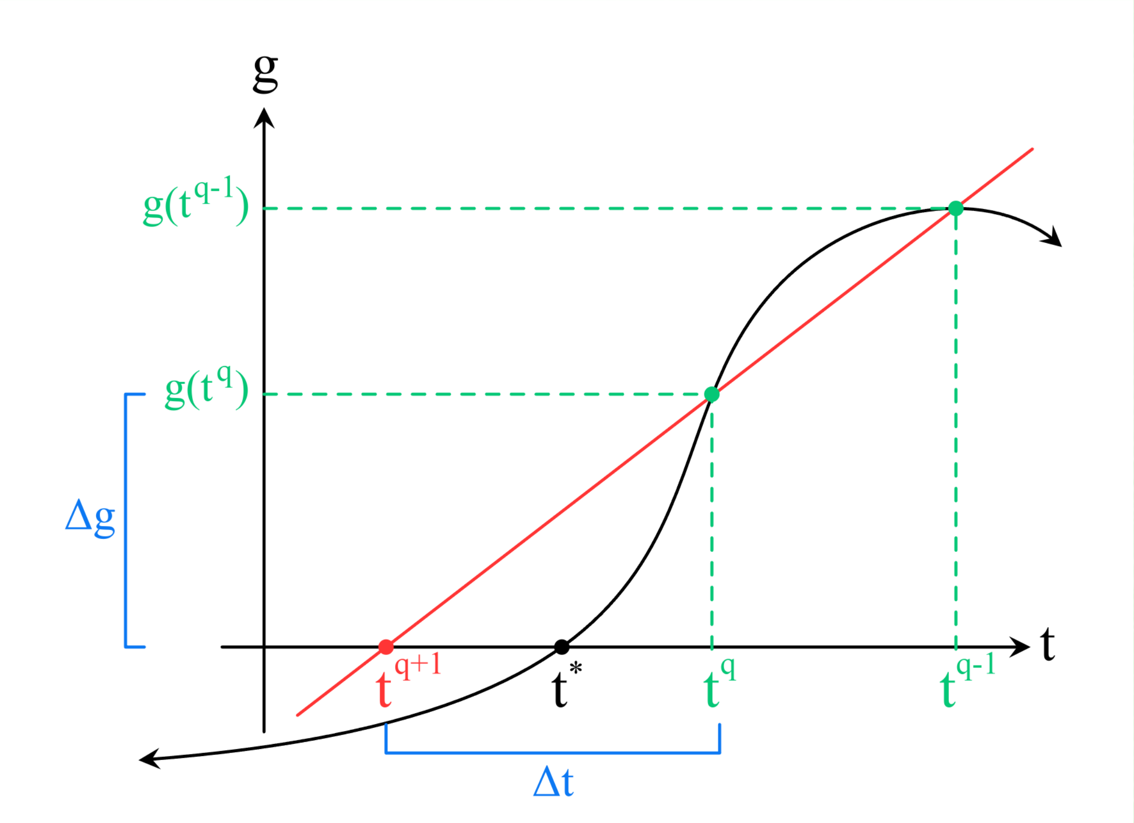
# Secant Method

- Replace $g'(t^q)$ in Newton's method with an estimate (a few choices for this)
- The standard approach 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 $p \approx 1.618$, when not degenerate
- Typically/often faster than Newton, since $g'$ is not needed and only a few extra iterations are required to obtain the same accuracy (for a reasonable 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

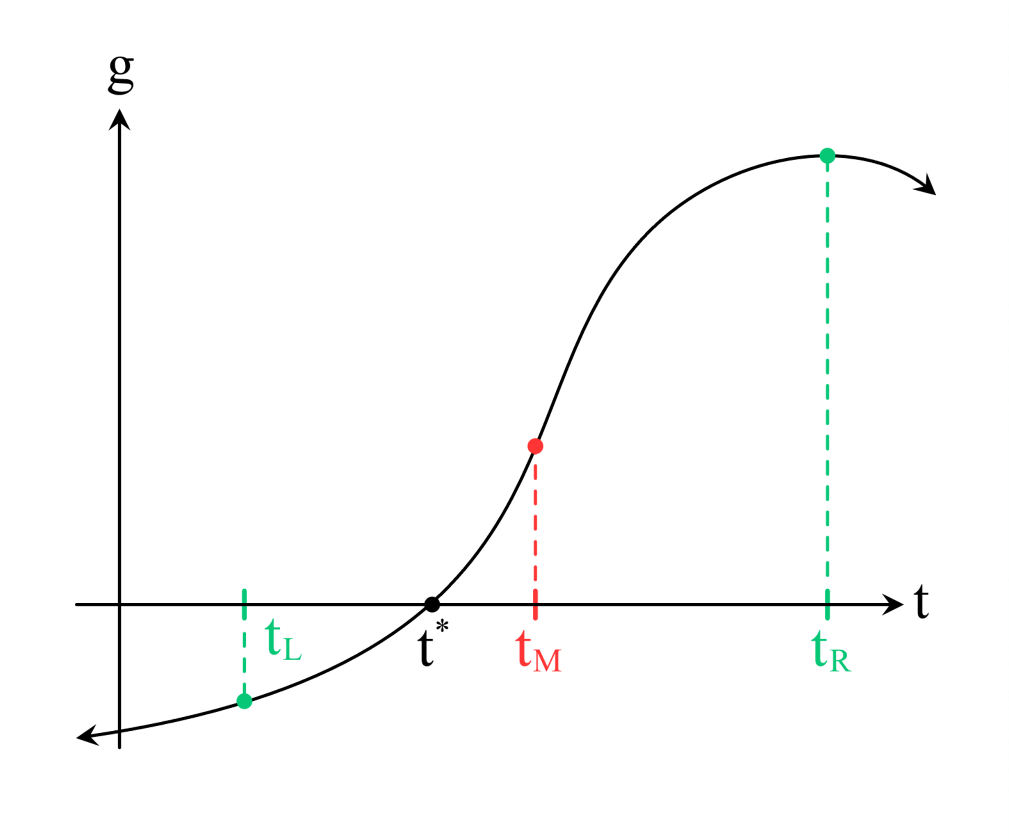
- If $g(t_L)g(t_R) < 0$, then (assuming continuity) the sign change indicates a root in the interval $[t_L, t_R]$
- Let $t_M = \frac{t_L + t_R}{2}$,
  - If $g(t_L)g(t_M) < 0$, set $t_R = t_M$
  - Otherwise, set $t_L = t_M$ knowing that $g(t_R)g(t_M) < 0$ is true
- Iterate until $t_R - t_L < \epsilon$

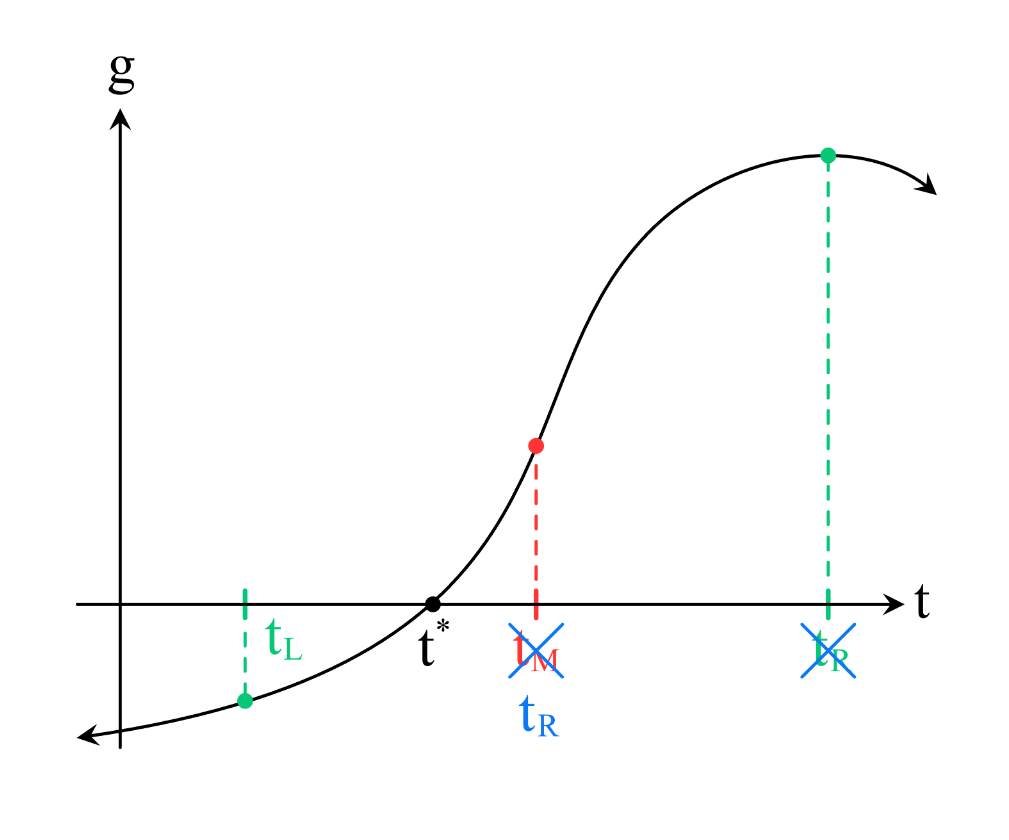
- Guaranteed to converge to a root in the interval (unlike Newton/Secant)

- 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)
- If Newton/Secant attempt to leave the current interval, instead use Bisection to continue shrinking the 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 $c^{q+1}(t) = c^q + t\Delta c^q$

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

- Scalar $\hat{f}(c^{q+1}(t))$ has system $J_{\hat{f}}^T\left(c^{q+1}(t)\right) = 0$ for critical points
- $\frac{\partial}{\partial t} J_{\hat{f}}^T\left(c^{q+1}(t)\right) = H_{\hat{f}}^T\left(c^{q+1}(t)\right)\Delta c^q$ and $\frac{\partial}{\partial t} J_{\hat{f}}\left(c^{q+1}(t)\right) = (\Delta c^q)^T H_{\hat{f}}\left(c^{q+1}(t)\right)$
  - $\frac{\partial}{\partial t} \left(J_{\hat{f}}^T\right)_i \left(c^{q+1}(t)\right) = \left(H_{\hat{f}}^T\right)_i \left(c^{q+1}(t)\right)\Delta c^q$

# Recall: Line Search (Unit 14)

- Given the linearization errors in $F'(c^q)\Delta c^q = (\beta - 1)F(c^q)$, the resulting $\Delta c^q$ can lead to a poor estimate for $c^{q+1}$ via $c^{q+1} = c^q + \Delta c^q$

- Instead, $\Delta c^q$ is often just used as a search direction, i.e. $c^{q+1} = c^q + \alpha^q \Delta c^q$
- The 1D (parameterized) line $c^{q+1}(\alpha) = c^q + \alpha \Delta c^q$ is the new domain
- Find an $\alpha$ with $F\left(c^{q+1}(\alpha)\right) = 0$ simultaneously for all equations

- <u>Safe Set</u> methods restrict $\alpha$ in various ways, e.g. $0 \leq \alpha \leq 1$

# Recall: Line Search (Unit 14)

- Since $F$ is vector valued, consider $g(\alpha) = F\left(c^{q+1}(\alpha)\right)^T F\left(c^{q+1}(\alpha)\right) = 0$
- Since $g(\alpha) \geq 0$, solutions to $F\left(c^{q+1}(\alpha)\right) = 0$ are minima of $g(\alpha)$
- $g(\alpha)$ might be strictly positive (with no $g(\alpha) = 0$), but minimizing $g(\alpha)$ might still help to make progress towards an $\alpha$ with $F\left(c^{q+1}(\alpha)\right) = 0$

- Option 1: find simultaneous roots of the vector valued $F\left(c^{q+1}(\alpha)\right) = 0$
- Option 2: find roots of or minimize $g(\alpha) = \frac{1}{2} F^T\left(c^{q+1}(\alpha)\right) F\left(c^{q+1}(\alpha)\right)$, to find or make progress towards an $\alpha$ with $F\left(c^{q+1}(\alpha)\right) = 0$

# Nonlinear Systems Problems

- Solve $J_F(c^q)\Delta c^q = (\beta - 1)F(c^q)$ for $\Delta c^q$ and use $c^{q+1}(t) = c^q + t\Delta c^q$ in $F\left(c^{q+1}(t)\right) = 0$
- <u>Option 1</u>: find simultaneous (for all $i$) roots for all the $g_i(t) = F_i\left(c^{q+1}(t)\right) = 0$
  - Here, $g_i'(t) = (J_F)_i\left(c^{q+1}(t)\right)\Delta c^q$
- <u>Option 2</u>: find roots of $g(t) = \frac{1}{2}F^T\left(c^{q+1}(t)\right)F\left(c^{q+1}(t)\right) = 0$
  - Here, $g'(t) = \frac{1}{2}F^T\left(c^{q+1}(t)\right)J_F\left(c^{q+1}(t)\right)\Delta c^q + \frac{1}{2}(\Delta c^q)^T J_F^T\left(c^{q+1}(t)\right)F\left(c^{q+1}(t)\right)$
  - Since both terms are scalars, $g'(t) = F^T\left(c^{q+1}(t)\right)J_F\left(c^{q+1}(t)\right)\Delta c^q$

# Recall: Optimization Problems (Unit 14)

- Minimize the scalar cost function $\hat{f}(c)$ by finding the critical points where $\nabla\hat{f}(c) = J_{\hat{f}}^T(c) = F(c) = 0$
- $F'(c^q)\Delta c^q = (\beta - 1)F(c^q)$ gives the search direction (as usual)
- Here, $F'(c) = J_F(c) = H_{\hat{f}}^T(c)$
- So, solve $H_{\hat{f}}^T(c^q)\Delta c^q = (\beta - 1)J_{\hat{f}}^T(c^q)$ to find the search direction $\Delta c^q$
- Option 1: find simultaneous roots of the vector valued $J_{\hat{f}}^T\left(c^{q+1}(\alpha)\right) = 0$, which are critical points of $\hat{f}(c)$
- Option 2: find roots of or minimize $g(\alpha) = \frac{1}{2}J_{\hat{f}}\left(c^{q+1}(\alpha)\right)J_{\hat{f}}^T\left(c^{q+1}(\alpha)\right)$, to find or make progress towards critical points of $\hat{f}(c)$
- Option 3: minimize $\hat{f}(c^{q+1}(\alpha))$ directly

# Optimization Problems

- Solve $H_{\hat{f}}^T(c^q)\Delta c^q = (\beta - 1)J_{\hat{f}}^T(c^q)$ for $\Delta c^q$ and use $c^{q+1}(t) = c^q + t\Delta c^q$ in $J_{\hat{f}}^T\left(c^{q+1}(t)\right) = 0$
- Option 1: find simultaneous (for all $i$) roots for all the $g_i(t) = (J_{\hat{f}}^T)_i\left(c^{q+1}(t)\right) = 0$ to find the critical points of $\hat{f}(c)$
  - Here, $g_i'(t) = (H_{\hat{f}}^T)_i\left(c^{q+1}(t)\right)\Delta c^q$
- Option 2: find roots of $g(t) = \frac{1}{2}J_{\hat{f}}\left(c^{q+1}(t)\right)J_{\hat{f}}^T\left(c^{q+1}(t)\right) = 0$ to find or make progress towards critical points of $\hat{f}(c)$
  - Here, $g'(t) = \frac{1}{2}J_{\hat{f}}\left(c^{q+1}(t)\right)H_{\hat{f}}^T\left(c^{q+1}(t)\right)\Delta c^q + \frac{1}{2}(\Delta c^q)^T H_{\hat{f}}\left(c^{q+1}(t)\right)J_{\hat{f}}^T\left(c^{q+1}(t)\right)$
  - Since both terms are scalars, $g'(t) = J_{\hat{f}}\left(c^{q+1}(t)\right)H_{\hat{f}}^T\left(c^{q+1}(t)\right)\Delta c^q$
- Option 3: minimize $\hat{f}(c^{q+1}(t))$ directly (see unit 16)