A Theory of Slack

How Economic Slack Shapes Markets, Business Cycles, and Policies

Pascal Michaillat

Draft version: May 2026
Draft URL: pascalmichaillat.org/18/

APPENDIX G.

Mathematical background

This final appendix lists background mathematical results that we use repeatedly throughout the book.[1] These results are well known. They are presented, proved, generalized, and discussed in other texts, so we do not rederive them but simply collect them here for convenience. A complete treatment of the results is provided in the following textbooks: Apostol (1967) for calculus; Boyd and Vandenberghe (2004) for convexity of functions and convex optimization; Acemoglu (2009) for optimal control; and Hirsch, Smale, and Devaney (2013) for differential equations and dynamical systems.

G.1. Miscellaneous results

RESULT G.1 (Intermediate-value theorem). *Consider a function f that is continuous on $[a, b]$. If $f(a)$ and $f(b)$ have opposite signs ($f(a)f(b) < 0$), then there exists $c \in (a, b)$ such that $f(c) = 0$. If, in addition, f is strictly monotone on $[a, b]$, then this zero is unique.*

RESULT G.2 (Euler's theorem for homogeneous functions). *Consider a differentiable function $f : (0, +\infty)^m \to (0, +\infty)$ that is homogeneous of degree $n > 0$, so that for all $x \in (0, +\infty)^m$ and $z > 0$:*

$$f(z \cdot x) = z^n \cdot f(x).$$

[1] The book assumes only basic familiarity with common functions (power, exponential, logarithmic, and so on), differential calculus (derivatives, linearization, implicit differentiation, and so on), and introductory probability and statistics (least-squares regressions, normal distribution, and so on). This appendix does not repeat these fundamentals; it collects results that are slightly more advanced.

Then the partial elasticities of the function add up to n:

$$\sum_{i=1}^{m} \epsilon_{x_i}^{f} = \sum_{i=1}^{m} \frac{x_i}{f(x)} \cdot \frac{\partial f}{\partial x_i} = n.$$

In the case of a function that has constant returns to scale (homogeneous of degree 1), the partial elasticities of f add up to 1:

$$\sum_{i=1}^{m} \epsilon_{x_i}^{f} = \sum_{i=1}^{m} \frac{x_i}{f(x)} \cdot \frac{\partial f}{\partial x_i} = 1.$$

RESULT G.3 (Poisson process). *A counting process $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if $N(0) = 0$, and for any $t > 0$ and $h > 0$, the increments $N(t+h) - N(t)$ are independent over disjoint time intervals and they are Poisson distributed with parameter λh:*

$$\mathbb{P}(N(t+h) - N(t) = k) = \frac{\exp(-\lambda h) \cdot (\lambda h)^k}{k!},$$

for any $k = 0, 1, 2, \ldots$. This implies that over a short time interval dt:

$$\mathbb{P}(N(t+dt) - N(t) = 1) = \lambda\, dt + o(dt), \qquad \mathbb{P}(N(t+dt) - N(t) > 1) = o(dt).$$

Thus, the probability that one event occurs in the near future is always just $\lambda\, dt$. Furthermore, the waiting time to the next event, T, has an exponential distribution with parameter λ:

$$\mathbb{P}(T < t) = 1 - \exp(-\lambda t), \qquad \mathbb{P}(T > t) = \exp(-\lambda t).$$

This implies in particular that the waiting time to the next event is memoryless:

$$\mathbb{P}(T > s + t \mid T > s) = \mathbb{P}(T > t).$$

It also implies that the mean waiting time simply is:

$$\mathbb{E}(T) = \frac{1}{\lambda}.$$

RESULT G.4 (Spectral identities). *For any square real matrix, the trace is equal to the sum of its eigenvalues and the determinant is equal to the product of its eigenvalues, counting repeated eigenvalues. If the matrix is symmetric, all its eigenvalues are real. In general, a real matrix may have complex eigenvalues, but nonreal eigenvalues occur in conjugate pairs, so trace and determinant remain real.*

RESULT G.5 (First-order recursions). *Consider the first-order recursion*

$$x_{t+1} = f(x_t).$$

A fixed point of the recursion x^* satisfies $x^* = f(x^*)$.

- *Convergence of the sequence $\{x(t)\}$ is not guaranteed. Depending on f and on the starting value, trajectories can converge, cycle, or diverge.*
- *If the sequence $\{x(t)\}$ is monotone and bounded, then it converges. Any limit point must be a fixed point x^*.*
- *If f is differentiable at x^*, there is local convergence if $|f'(x^*)| < 1$ and local divergence if $|f'(x^*)| > 1$.*

RESULT G.6 (Leibniz's integral rule). *Consider the integral*

$$I(z) = \int_{a(z)}^{b(z)} f(x, z)\, dx,$$

where x is the integration variable, z is a parameter, the functions a and b are differentiable, and the function f is continuously differentiable. Then the effect of a change in the parameter on the integral is given by:

$$\frac{dI}{dz} = \int_{a(z)}^{b(z)} \frac{\partial f}{\partial z}\, dx + \frac{db}{dz} f(b(z), z) - \frac{da}{dz} f(a(z), z).$$

RESULT G.7 (Implicit function theorem). *Consider a point (x_0, y_0) that satisfies $f(x_0, y_0) = 0$, where the function f is continuously differentiable in a neighborhood of (x_0, y_0) and $\partial f/\partial y(x_0, y_0) \neq 0$. Then there exists a neighborhood of x_0 in which there is a unique differentiable function $y(x)$ such that $y(x_0) = y_0$ and $f(x, y(x)) = 0$. In that neighborhood, the derivative of y with respect to x satisfies*

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dx} = 0.$$

In other words, the derivative of y is determined by the partial derivatives of f:

$$\frac{dy}{dx} = -\frac{\partial f/\partial x}{\partial f/\partial y}.$$

G.2. The exponential and logarithm functions

RESULT G.8. *The exponential function is strictly convex, so it is above all its tangents, and in particular above the tangent at $x = 0$:*

$$\exp(x) \geq 1 + x$$

for all $x \in \mathbb{R}$, with equality only at $x = 0$.

RESULT G.9. *The logarithm function is strictly concave, so it is below all its tangents, and in particular below the tangent at $x = 1$:*

$$\ln(x) \leq x - 1,$$

for all $x > 0$, with equality only at $x = 1$.

RESULT G.10. *The exponential function equals its Taylor series at 0 (Maclaurin series):*

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

for all $x \in \mathbb{R}$. This is an exact identity, not a local approximation: because the exponential function is analytic on $\mathbb{R}$, this series converges to the function everywhere.

G.3. Rational functions

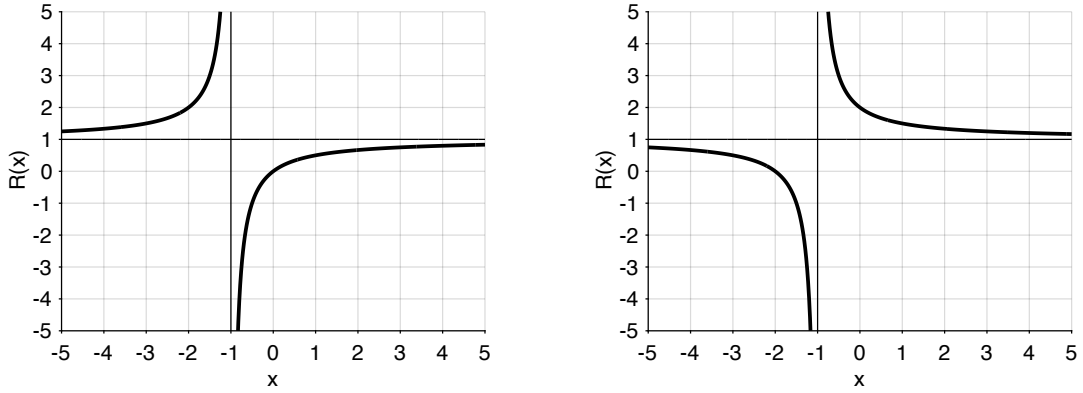
RESULT G.11. *Consider a linear-over-linear rational function:*

$$R(x) = \frac{ax + b}{cx + d},$$

with $ad - bc \neq 0$ and $c \neq 0$. The function admits a single pole at $x = -d/c$, where $R(x) \to \pm\infty$, and a single asymptote at $y = a/c$, when $x \to \pm\infty$. If $a \neq 0$, the function admits a single root at $x = -b/a$, where $R(x) = 0$. If $a = 0$, the function has no roots and its asymptote is $y = 0$. If $ad - bc > 0$, the function is strictly increasing and strictly convex on $(-\infty, -d/c)$ and strictly increasing and strictly concave on $(-d/c, +\infty)$ (as illustrated in figure G.1A). If $ad - bc < 0$, the function is strictly decreasing and strictly concave on $(-\infty, -d/c)$ and strictly decreasing and strictly convex on $(-d/c, +\infty)$ (as illustrated in figure G.1B).

RESULT G.12. *Consider the generic rational function*

$$R(x) = \frac{P(x)}{Q(x)},$$



A. First type: $ad - bc > 0$ B. Second type: $ad - bc < 0$

FIGURE G.1. The two types of linear-over-linear rational functions

Panel A plots $y = x/(x+1)$. Panel B plots $y = (x+2)/(x+1)$. For both, the pole is $x = -1$ and the asymptote is $y = 1$.

where P and Q are polynomials with no common factor. Many of the properties of the function can be determined by graphing it:

- *Each real root of Q is a pole of R, where it presents a vertical asymptote.*
- *Each real root of P is a root of R, where it comes in contact with the x-axis:*
  - *If the root has multiplicity m and m is odd, the graph crosses the x-axis.*
  - *If the root has multiplicity m and m is even, the graph touches the x-axis.*
- *The limits of R at infinity are determined by polynomial degrees:*
  - *If* $\deg P < \deg Q$*, then* $\lim_{x \to \pm\infty} R(x) = 0$.
  - *If* $\deg P = \deg Q$*, the limit is the ratio of leading coefficients.*
  - *If* $\deg P = \deg Q + 1$*, there is an oblique asymptote obtained by polynomial division.*
  - *If* $\deg P > \deg Q + 1$*, there is a polynomial asymptote obtained by polynomial division.*

G.4. Convexity of functions

The results below are adapted from Boyd and Vandenberghe (2004, chapter 3).

RESULT G.13. *The function f is concave iff the function $-f$ is convex.*

From this result, we can infer the concavity results below from the convexity results. Nevertheless, to improve user-friendliness, we state both convex and concave cases.

RESULT G.14. *For a twice-differentiable function of one variable, $f(x)$, defined on a convex set, concavity and convexity can be checked from its second derivative:*

- *The function is concave if* $f''(x) \leq 0$ *on the set.*
- *The function is strictly concave if* $f''(x) < 0$ *on the set.*
- *The function is convex if* $f''(x) \geq 0$ *on the set.*
- *The function is strictly convex if* $f''(x) > 0$ *on the set.*

RESULT G.15. *For a twice-differentiable function of two variables,* $f(x, y)$*, defined on a convex set, concavity and convexity can be checked from its Hessian:*

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$

- *The function is concave if the Hessian is negative semidefinite on the set. This requires that the Hessian's two, real eigenvalues are negative, and in turn that its determinant is positive and its trace is negative.*
- *The function is strictly concave if the Hessian is negative definite. This requires that the Hessian's two, real eigenvalues are strictly negative, and in turn that its determinant is strictly positive and its trace is strictly negative.*
- *The function is convex if the Hessian is positive semidefinite on the set. This requires that the Hessian's two, real eigenvalues are positive, and in turn that its determinant is positive and its trace is positive.*
- *The function is strictly convex if the Hessian is positive definite. This requires that the Hessian's two, real eigenvalues are strictly positive, and in turn that its determinant is strictly positive and its trace is strictly positive.*

RESULT G.16. *Consider an affine function* $g(x) = ax + b$ *with* $a \neq 0$*.*

- *If* f *is (strictly) convex, then* $f \circ g$ *is (strictly) convex.*
- *If* f *is (strictly) concave, then* $f \circ g$ *is (strictly) concave.*
- *If* f *is convex and* $a \geq 0$*, then* $g \circ f$ *is convex.*
- *If* f *is concave and* $a \geq 0$*, then* $g \circ f$ *is concave.*
- *If* f *is strictly convex and* $a > 0$*, then* $g \circ f$ *is strictly convex.*
- *If* f *is strictly concave and* $a > 0$*, then* $g \circ f$ *is strictly concave.*

RESULT G.17. *The positive weighted sum of convex functions is convex. If any term in the sum is a strictly convex function with a strictly positive weight, the sum is strictly convex. Similarly, the positive weighted sum of concave functions is concave. If any term in the sum is a strictly*

concave function with a strictly positive weight, the sum is strictly concave.

RESULT G.18. *Consider first two twice-differentiable functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. The composition $f \circ g$ of the two functions has the following properties:*

- *If f is convex and increasing, and g is convex, then $f \circ g$ is convex.*
- *If f is convex and decreasing, and g is concave, then $f \circ g$ is convex.*
- *If f is concave and increasing, and g is concave, then $f \circ g$ is concave.*
- *If f is concave and decreasing, and g is convex, then $f \circ g$ is concave.*

If the two functions are defined not on $\mathbb{R}$ but on some convex subset of $\mathbb{R}$, the results continue to hold once we add a constraint on the domain of f:

- *If f is convex, f is increasing on an interval $(-\infty, a)$, and g is convex, then $f \circ g$ is convex.*
- *If f is convex, f is decreasing on an interval $(a, +\infty)$, and g is concave, then $f \circ g$ is convex.*
- *If f is concave, f is increasing on an interval $(a, +\infty)$, and g is concave, then $f \circ g$ is concave.*
- *If f is concave, f is decreasing on an interval $(-\infty, a)$, and g is convex, then $f \circ g$ is concave.*

RESULT G.19. *Consider a twice-differentiable bijection f. Then its inverse f^{-1} has the following properties:*

- *If f is strictly increasing and strictly convex, f^{-1} is strictly increasing and strictly concave.*
- *If f is strictly decreasing and strictly convex, f^{-1} is strictly decreasing and strictly convex.*
- *If f is strictly increasing and strictly concave, f^{-1} is strictly increasing and strictly convex.*
- *If f is strictly decreasing and strictly concave, f^{-1} is strictly decreasing and strictly concave.*

G.5. Convex optimization

Convex optimization is concerned with static optimization of convex (or concave) functions over convex sets. The results below are adapted from Boyd and Vandenberghe (2004, chapter 4).

RESULT G.20. *Consider the convex optimization problem:*

$$\min_{x \in \mathcal{X}} f(x) \tag{G.1}$$

If the objective function f is differentiable and convex and the feasible set $\mathcal{X} \subset \mathbb{R}^n$ is convex, then any interior point satisfying the n first-order conditions

$$\frac{\partial f}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n,$$

is a global minimum. Furthermore, if the objective function is strictly convex, the global minimum is unique.

RESULT G.21. *Consider the convex optimization problem:*

$$\max_{x \in \mathcal{X}} f(x) \tag{G.2}$$

If the objective function f is differentiable and concave and the feasible set $\mathcal{X} \subset \mathbb{R}^n$ is convex, then any interior point satisfying the n first-order conditions

$$\frac{\partial f}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n,$$

is a global maximum. Furthermore, if the objective function is strictly concave, the global maximum is unique.

RESULT G.22. *Consider the feasible set of an optimization problem described by m inequality constraints:*

$$f_i(x) \leq 0, \quad i = 1, 2, \ldots, m.$$

If the inequality-constraint functions f_i are convex, then the feasible set of the optimization problem is convex. For any interior point of the feasible set, the inequalities are not binding: $f_i(x) < 0$ for all i.

RESULT G.23. *If the optimization problem contains equality constraints, it is often convenient to eliminate them by substituting the appropriate number of variables out of the objective $f(x_1, x_2, \ldots, x_n)$. When this is possible, only inequality constraints are left after substitution, and the optimization problem can be written as in* (G.1) *or* (G.2)*. The substitution approach generally works if the equality constraints are affine (of the form $\sum_{i=1}^{n} a_{i,k} x_i = b_k$) or take certain simple nonlinear forms. If the substitution approach is not available, or not convenient, one typically keeps the equality constraints and uses a Lagrangian approach.*

G.6. Optimal control

Optimal control is concerned with dynamic optimization in continuous time. The results in this section are adapted from Acemoglu (2009, chapter 7).

RESULT G.24. *Consider the optimal control problem with control variable $v_1(t), v_2(t), \ldots, v_n(t)$ and state variable $w_1(t), w_2(t), \ldots, w_m(t)$:*

$$\max_{v_1(t), v_2(t), \ldots, v_n(t), t \geq 0} \int_0^{\infty} e^{-\rho t} f(v_1(t), v_2(t), \ldots, w_1(t), w_2(t), \ldots, t)\, dt \tag{G.3}$$

subject to the m laws of motion

$$\dot{w}_i(t) = g_i(v_1(t), v_2(t), \ldots, w_1(t), w_2(t), \ldots, t), \quad i = 1, 2, \ldots, m,$$

and given initial conditions and admissibility constraints. The discount rate ρ *is strictly positive, and the functions* f *and* g_i *are twice continuously differentiable in state and control arguments, and continuous in time. Then the current-value Hamiltonian is*

$$\mathcal{H}(t) = f(v_1(t), \ldots, w_1(t), \ldots, t) + \sum_{i=1}^{m} q_i(t) \cdot g_i(v_1(t), \ldots, w_1(t), \ldots, t),$$

where $q_1(t), q_2(t), \ldots, q_m(t)$ *are the costate variables for the state variables. Just like in result G.23, we assume here that all equality constraints have been substituted out of the optimization problem. If the substitution is not possible or convenient, extra Lagrangian terms must be added to the Hamiltonian.*

RESULT G.25. *For the optimal control problem* (G.3)*, the necessary conditions for an interior maximum are the Pontryagin first-order conditions:*

$$\frac{\partial \mathcal{H}}{\partial v_j} = 0, \quad j = 1, 2, \ldots, n \tag{G.4}$$

$$\frac{\partial \mathcal{H}}{\partial w_i} = \rho q_i(t) - \dot{q}_i(t), \quad i = 1, 2, \ldots, m. \tag{G.5}$$

In addition, appropriate transversality conditions rule out explosive paths that violate optimality. A typical transversality condition is

$$\lim_{t \to \infty} \exp(-\rho t) q_i(t) w_i(t) = 0.$$

The transversality condition rules out asymptotically wasteful paths by requiring that the discounted shadow value of remaining state variables vanishes as $t \to \infty$.

RESULT G.26. *Consider the optimal control problem* (G.3)*, and assume that the laws of motion of the state variables are affine:*

$$\dot{\mathbf{w}}(t) = \mathbf{A}(t)\mathbf{w}(t) + \mathbf{B}(t)\mathbf{v}(t) + \mathbf{u}(t),$$

where $\mathbf{w}(t) = [w_1(t), \ldots, w_m(t)] \in \mathbb{R}^m$ *is the vector of state variables,* $\mathbf{v}(t) = [v_1(t), \ldots, v_n(t)] \in \mathbb{R}^n$ *is the vector of control variables, and* $\mathbf{u}(t) = [u_1(t), \ldots, u_m(t)] \in \mathbb{R}^m$ *is a vector of forcing terms in the laws of motion. Suppose that:*

- *The set of admissible controls and states is convex.*
- *For each* t*, the objective function* $f(\mathbf{v}, \mathbf{w}, t)$ *is concave in* $(\mathbf{v}, \mathbf{w})$.

- *Transversality conditions are satisfied.*

An admissible interior path is any path for states and controls that satisfies the initial conditions and laws of motion, stays strictly inside the feasible set at every date, and satisfies any additional admissibility conditions (such as no-Ponzi conditions). Then any admissible interior path $(\mathbf{v}(t), \mathbf{w}(t))$ *that satisfies the Pontryagin first-order conditions is a global maximum. If the objective function is strictly concave, the optimum is unique.*

RESULT G.27. *Consider the optimal control problem* (G.3)*. Suppose that:*

- *The set of admissible controls and states is convex.*
- *For each* t*, and along the relevant costate path, the Hamiltonian* $\mathcal{H}(\mathbf{v}, \mathbf{w}, \mathbf{q}(t), t)$ *is concave in* $(\mathbf{v}, \mathbf{w})$*.*
- *Transversality conditions are satisfied.*

Then any admissible interior path that satisfies the Pontryagin first-order conditions is a global maximum. If the Hamiltonian is strictly concave in $(\mathbf{v}, \mathbf{w})$*, the optimum is unique.*

G.7. Differential equations

Differential equations are functional equations in continuous time.

RESULT G.28. *The exponential function is defined as the only solution to the differential equation*

$$\dot{x}(t) - x(t) = 0,$$

with initial condition $x(0) = 1$*. Hence, the only solution to the autonomous and homogeneous linear first-order differential equation*

$$\dot{x}(t) - \lambda x(t) = 0,$$

with initial condition $x(t_0) = x_0$*, is*

$$x(t) = x_0 \exp(\lambda (t - t_0)).$$

Moreover, the only solution to the autonomous linear first-order differential equation

$$\dot{x}(t) - \lambda x(t) = \phi,$$

with initial condition $x(t_0) = x_0$*, is*

$$x(t) = \frac{\phi}{\lambda} + \left(x_0 - \frac{\phi}{\lambda}\right) \exp(\lambda (t - t_0)).$$

RESULT G.29. *The only solution to the generic linear first-order differential equation*

$$\dot{x}(t) - \lambda(t)x(t) = \phi(t),$$

with initial condition $x(t_0) = x_0$*, is*

$$x(t) = x_0 \exp\left(\int_{t_0}^{t} \lambda(s)\, ds\right) + \int_{t_0}^{t} \phi(z) \exp\left(\int_{z}^{t} \lambda(s)\, ds\right) dz.$$

G.8. Dynamical systems

Dynamical systems are collections of differential equations. The results in this section are adapted from Hirsch, Smale, and Devaney (2013, chapters 1–6).

G.8.1. Linear dynamical system

We consider a 2×2 homogeneous autonomous linear first-order dynamical system:

$$\begin{aligned} \dot{x}(t) &= a \cdot x(t) + b \cdot y(t) \\ \dot{y}(t) &= c \cdot x(t) + d \cdot y(t). \end{aligned}$$

The matrix governing the linear system is

$$\mathbf{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{G.6}$$

We denote by $\lambda_1 \in \mathbb{C}$ and $\lambda_2 \in \mathbb{C}$ the two eigenvalues of $\mathbf{M}$, assumed to be nonzero and distinct. We denote by $\mathbf{v}_1 \in \mathbb{C}^2$ and $\mathbf{v}_2 \in \mathbb{C}^2$ the linearly independent eigenvectors associated with the eigenvalues λ_1 and λ_2.

RESULT G.30. *Assume that* λ_1 *and* λ_2 *are real. Without loss of generality, we assume* $\lambda_1 < \lambda_2$*. Then the solution to the linear system takes the form*

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \alpha_1 \exp(\lambda_1 t)\mathbf{v}_1 + \alpha_2 \exp(\lambda_2 t)\mathbf{v}_2, \tag{G.7}$$

where $\alpha_1 \in \mathbb{R}$ *and* $\alpha_2 \in \mathbb{R}$ *are constants determined by boundary conditions. The system exhibits nodal dynamics:*

- *The system is a source when* $\lambda_1 > 0$ *and* $\lambda_2 > 0$*. The solutions are tangent to* $\mathbf{v}_1$ *when* $t \to -\infty$ *and are parallel to* $\mathbf{v}_2$ *when* $t \to +\infty$*.*

- *The system is a sink when* $\lambda_1 < 0$ *and* $\lambda_2 < 0$*. The solutions are tangent to* $\mathbf{v}_1$ *when* $t \to -\infty$ *and are parallel to* $\mathbf{v}_2$ *when* $t \to +\infty$*.*
- *The system is a saddle when* $\lambda_1 < 0$ *and* $\lambda_2 > 0$*. The vector* $\mathbf{v}_1$ *gives the direction of the stable line (saddle path) while the vector* $\mathbf{v}_2$ *gives the direction of the unstable line.*

In that way, in all cases, trajectories move from the $\mathbf{v}_1$*-direction to the* $\mathbf{v}_2$*-direction as time increases.*

RESULT G.31. *Assume that* λ_1 *and* λ_2 *are complex conjugates. We write the eigenvalues as* $\lambda_1 = \theta + i\vartheta$ *and* $\lambda_2 = \theta - i\vartheta$*. We also write the eigenvector associated with* λ_1 *as* $\mathbf{v}_1 + i\mathbf{v}_2$*, where the vectors* $\mathbf{v}_1 \in \mathbb{R}^2$ *and* $\mathbf{v}_2 \in \mathbb{R}^2$ *are linearly independent. Then the solution to the linear system takes the form*

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \exp(\theta t)\,[\mathbf{v}_1, \mathbf{v}_2] \begin{bmatrix} \cos(\vartheta t) & \sin(\vartheta t) \\ -\sin(\vartheta t) & \cos(\vartheta t) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix},$$

where $[\mathbf{v}_1, \mathbf{v}_2] \in \mathbb{R}^{2\times 2}$ *is a* 2×2 *matrix, and* $\alpha_1 \in \mathbb{R}$ *and* $\alpha_2 \in \mathbb{R}$ *are constants determined by boundary conditions. The system exhibits spiral dynamics:*

- *The system is a spiral source if* $\theta > 0$*. Then the solutions wind periodically around the origin, moving away from it.*
- *The system is a spiral sink if* $\theta < 0$*. Then the solutions wind periodically around the origin, moving toward it.*
- *The system is a center if* $\theta = 0$*. Then the solutions circle around the origin.*

RESULT G.32. *The discriminant of the matrix* $\mathbf{M}$*,*

$$\Delta = \operatorname{tr}(\mathbf{M})^2 - 4\det(\mathbf{M}), \tag{G.8}$$

distinguishes the cases with real and complex eigenvalues:

- *If* $\Delta > 0$*, the eigenvalues are real and the system exhibits nodal dynamics.*
- *If* $\Delta < 0$*, the eigenvalues are complex and the system exhibits spiral or center dynamics.*

The trace and determinant of the matrix $\mathbf{M}$ *further help classify the system dynamics:*

- *If* $\det(\mathbf{M}) < 0$*, eigenvalues are real and have opposite signs, so the system is a saddle.*
- *If* $\det(\mathbf{M}) > 0$ *and* $\operatorname{tr}(\mathbf{M}) < 0$*, both eigenvalues have negative real parts, so the system is a sink.*
- *If* $\det(\mathbf{M}) > 0$ *and* $\operatorname{tr}(\mathbf{M}) > 0$*, both eigenvalues have positive real parts, so the system is a source.*

- *If* $\det(\mathbf{M}) > 0$ *and* $\operatorname{tr}(\mathbf{M}) = 0$*, both eigenvalues are purely imaginary, so the system is a center.*

The dynamics are degenerate in all other cases:

- *If* $\Delta = 0$*, the eigenvalues are real and repeated, so the system exhibits degenerate nodal dynamics.*
- *If* $\det(\mathbf{M}) = 0$*, at least one of the eigenvalues is 0, so the system exhibits degenerate dynamics.*

RESULT G.33. *A nonhomogeneous autonomous linear first-order dynamical system can be solved using the results on homogeneous systems. Consider the nonhomogeneous system* $\dot{\mathbf{v}}(t) = \mathbf{M}\mathbf{v}(t) - \mathbf{u}$*, where* $\mathbf{v}(t) \in \mathbb{R}^2$*,* $\mathbf{M} \in \mathbb{R}^{2\times 2}$ *with* $\det \mathbf{M} \neq 0$*, and* $\mathbf{u} \in \mathbb{R}^2$ *with* $\mathbf{u} \neq 0$*. The equilibrium point of the system is* $\mathbf{v}^* = \mathbf{M}^{-1}\mathbf{u}$*. Solving the nonhomogeneous system is equivalent to solving the homogeneous system* $\dot{\mathbf{w}}(t) = \mathbf{M}\mathbf{w}(t)$ *and then recovering the variable* $\mathbf{v}(t) = \mathbf{w}(t) + \mathbf{v}^*$*.*

G.8.2. Phase diagrams

Without solving for eigenvalues and eigenvectors explicitly, we can study the properties of a 2×2 autonomous linear first-order dynamical system by drawing its phase diagram.

Here we construct the phase diagram for a nonhomogeneous dynamical system:

$$\dot{x}(t) = a \cdot x(t) + b \cdot y(t) + w \tag{G.9}$$

$$\dot{y}(t) = c \cdot x(t) + d \cdot y(t) + z, \tag{G.10}$$

Let $\mathbf{M}$ be the matrix associated with the system, given by (G.6). We assume that the eigenvalues of the system are nonzero, real, and distinct (so the discriminant of $\mathbf{M}$ is strictly positive and its determinant is nonzero). This implies that the dynamics of the system are nodal and nondegenerate.

Drawing the phase diagram of a two-variable system is useful to understand the main features of the dynamical system without solving for $x(t)$ and $y(t)$ explicitly. Figure G.2 illustrates the construction of the phase diagram with six panels: G.2A–G.2B for the saddle case, G.2C–G.2D for the source case, and G.2E–G.2F for the sink case.

We first plot the nullclines, which are the loci $\dot{x} = 0$ and $\dot{y} = 0$. The locus for $\dot{x} = 0$ is a straight line given by

$$ax + by + w = 0. \tag{G.11}$$

The locus for $\dot{y} = 0$ is a straight line given by

$$cx + dy + z = 0. \tag{G.12}$$

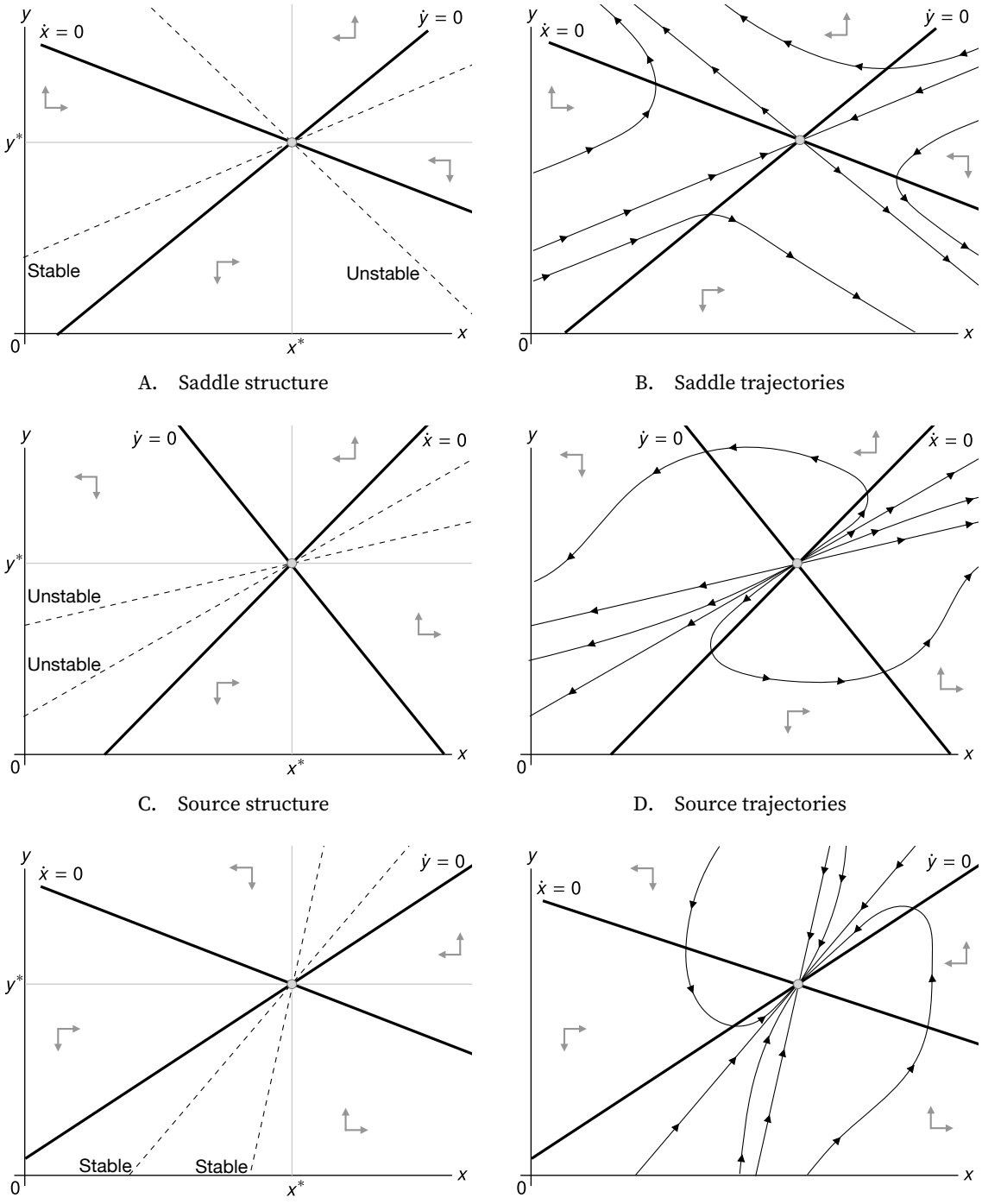


FIGURE G.2. Phase diagram for a 2 × 2 linear dynamical system

In these illustrative phase diagrams, we assume that the eigenvalues of the system are real and distinct and nonzero, so the dynamics are nodal and nondegenerate.

Next we place the equilibrium point, (x^*, y^*). The equilibrium is given by the intersection of the two nullclines:

$$x^* = \frac{bz - dw}{ad - bc},$$
$$y^* = \frac{cw - az}{ad - bc}.$$

These two nullclines partition the plane into four regions where the signs of $\dot{x}$ and $\dot{y}$ are constant. In each quadrant, the solutions have specific directions: going up or down and going left or right. We describe the direction field with directional arrows in the diagram. These arrows determine the direction of the solution's trajectories over time anywhere on the phase diagram. For example, from (G.9), we see that any point above the $\dot{x} = 0$ line (given by (G.11)) has $\dot{x} > 0$ if $b > 0$ and $\dot{x} < 0$ if $b < 0$. Conversely, any point below the $\dot{x} = 0$ line has $\dot{x} < 0$ if $b > 0$ and $\dot{x} > 0$ if $b < 0$. We represent these properties by a horizontal arrow pointing east for any point with $\dot{x} > 0$ and a horizontal arrow pointing west for any point with $\dot{x} < 0$.

Similarly, from (G.10), we see that any point above the $\dot{y} = 0$ line (given by (G.12)) has $\dot{y} > 0$ if $d > 0$ and $\dot{y} < 0$ if $d < 0$. Conversely, any point below the $\dot{y} = 0$ line has $\dot{y} < 0$ if $d > 0$ and $\dot{y} > 0$ if $d < 0$. We represent these properties by a vertical arrow pointing north for any point with $\dot{y} > 0$ and a vertical arrow pointing south for any point with $\dot{y} < 0$.

We then draw the stable or unstable lines through the equilibrium point. These lines, corresponding to the eigenvectors of $\mathbf{M}$, organize all trajectories. Indeed, the stable and unstable lines act as separatrices that cannot be crossed: any solution remains in a region of the plane delimited by the lines.

Finally, using the directional arrows and stable and unstable lines, we can draw trajectories that satisfy the system of differential equations. These are solutions to the system. To select a specific solution among all possible solutions, we would need to specify an initial condition.

If the system is a saddle, the stable line is the set of initial conditions that converge to the equilibrium as $t \to +\infty$, while the unstable line is the set that converges to the equilibrium as $t \to -\infty$. Trajectories are tangent to the stable line as $t \to -\infty$ and tangent to the unstable line as $t \to +\infty$.

If the system is a source, trajectories move away from equilibrium in forward time. All the trajectories are tangent to one unstable direction as $t \to -\infty$ (the one associated with the smallest eigenvalue) and tangent to the other unstable direction as $t \to +\infty$ (the one associated with the largest eigenvalue).

If the system is a sink, trajectories move toward equilibrium in forward time. Again, all the trajectories are tangent to one unstable direction as $t \to -\infty$ (the one associated with the lowest eigenvalue) and tangent to the other unstable direction as $t \to +\infty$ (the

one associated with the highest eigenvalue).

G.8.3. Nonlinear dynamical system

The methodology that we developed for 2×2 linear dynamical systems is also useful to study 2×2 nonlinear dynamical systems.

Consider the autonomous nonlinear system

$$\begin{aligned} \dot{x}(t) &= f(x(t), y(t)) \\ \dot{y}(t) &= g(x(t), y(t)). \end{aligned}$$

An equilibrium (x^*, y^*) is defined by

$$f(x^*, y^*) = 0, \qquad g(x^*, y^*) = 0.$$

To study local dynamics around (x^*, y^*), define deviations

$$u(t) = x(t) - x^*, \qquad v(t) = y(t) - y^*.$$

A first-order linearization around the equilibrium gives

$$\begin{bmatrix} \dot{u}(t) \\ \dot{v}(t) \end{bmatrix} = J(x^*, y^*) \begin{bmatrix} u(t) \\ v(t) \end{bmatrix},$$

where the Jacobian matrix at (x^*, y^*) is defined by

$$J(x^*, y^*) = \begin{bmatrix} \frac{\partial f}{\partial x}(x^*, y^*) & \frac{\partial f}{\partial y}(x^*, y^*) \\ \frac{\partial g}{\partial x}(x^*, y^*) & \frac{\partial g}{\partial y}(x^*, y^*) \end{bmatrix}.$$

Therefore, the local behavior of the nonlinear system near the equilibrium (x^*, y^*) is determined by the Jacobian $J(x^*, y^*)$, as long as the equilibrium is hyperbolic (no eigenvalue with zero real part). In particular, results G.30 and G.31 describe the solutions $(u(t), v(t))$ of the linearized system based on the eigenvalues and eigenvectors of the Jacobian. The signs of the discriminant, trace, and determinant of the Jacobian classify the equilibrium locally as saddle, sink, or source, as described by result G.32. It is also possible to build a phase diagram using the Jacobian, just as in the linear case described in figure G.2. The phase diagram provides a local approximation of the nonlinear system around the equilibrium.