Differentiation in $\mathbb{R}^n$

Robert J. Carroll

Last revised: 6 May 2026

---

1 Motivation

Comparative statics is the central methodological move of formal political-economy theory. A modeler asks: how does the optimal candidate position shift when voter ideal points move? How does the equilibrium tax rate change when the cost of public goods rises? How does a coalition's optimal allocation respond to a budget shock? How does a Bayesian updater's posterior shift when the prior changes? The answers to these questions are derivatives. The candidate's position is a function of the voter distribution; the equilibrium tax is a function of the cost parameter; the coalition's allocation is a function of the budget; the posterior is a function of the prior. To say how the dependent variable shifts when the parameter shifts is to compute a partial derivative, evaluate it at the relevant point, and read off the sign and magnitude.

In a one-dimensional setting the calculus a reader brings from undergraduate coursework is enough. Most political-economy applications, however, are multidimensional. The candidate's position is a vector in $\mathbb{R}^k$. The voter distribution is a function on $\mathbb{R}^k$. The equilibrium tax is part of a system of equilibrium conditions involving multiple variables and parameters. The objects that play the role of "derivative" in $\mathbb{R}^n$ — the gradient, the Jacobian, the Hessian — are the multivariable extensions of the single-variable derivative, and this handout introduces them with the apparatus needed to compute, manipulate, and interpret them.

Three structural results carry the load. The *chain rule* (§3) tells us how to compose derivatives: when the dependent variable depends on intermediate variables that themselves depend on parameters, the marginal effect decomposes via the Jacobian. *Taylor's theorem* (§5) localizes a smooth function around a point: any twice-differentiable function looks like a quadratic form locally, and the second-order behavior is governed by the Hessian. The *implicit function theorem* (§6) is the comparative-statics workhorse: when the equilibrium is defined implicitly by a system of equations, the theorem gives the derivative of the equilibrium with respect to its parameters without having to solve for the equilibrium first. Most of the comparative-statics arguments in formal political-economy theory are implicit-function-theorem applications, sometimes paired with the envelope theorem from the next handout.

The handout sits as the lead-in to the optimization cluster. The next handout takes up convex sets and concave functions, where the second-order machinery of §4 acquires its diagnostic role. The handout after that does static optimization (FOC, SOC, Lagrangians, KKT, envelope), and the dynamic-optimization handout closes the cluster.

2 Partial derivatives and gradients

A function of multiple variables changes in different ways depending on which variable is moving. A candidate's vote share might increase with persuasion effort and decrease with persuasion-spending visibility, and the modeler wants to track each direction's effect separately. A voter's utility function on a multidimensional policy space has different sensitivities to different policy dimensions, and the modeler wants to know which dimension's change has the biggest effect. The right tool is the *partial derivative* — the derivative with respect to one variable holding the others fixed — and the *gradient*, which packages all of the partial derivatives into a single vector pointing in the direction of steepest increase.

Definition 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function, and let $\boldsymbol{x} \in \mathbb{R}^n$. The *partial derivative* of f with respect to x_i at $\boldsymbol{x}$ is

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x})}{h},$$

when the limit exists, where $\boldsymbol{e}_i$ is the ith standard basis vector.

The partial derivative is the ordinary single-variable derivative of the function obtained by holding all variables except x_i fixed. So the techniques and theorems for single-variable derivatives apply directly to each partial.

Definition 2. The *gradient* of $f : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x}$ is the vector of partial derivatives

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \dots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \right) \in \mathbb{R}^n.$$

f is *continuously differentiable* at $\boldsymbol{x}$, written $f \in C^1$ at $\boldsymbol{x}$, if every partial $\frac{\partial f}{\partial x_i}$ is continuous in a neighborhood of $\boldsymbol{x}$. $f \in C^1$ on U if $f \in C^1$ at every $\boldsymbol{x} \in U$.

The gradient is a vector — specifically, an element of $\mathbb{R}^n$ whose ith coordinate is the marginal effect of increasing x_i. Two structural readings are central to political-economy applications, and both rely on the inner-product machinery of §6 of the linear-algebra cluster.

Proposition 3 (Directional derivative). *If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at $\boldsymbol{x}$, then for any unit vector $\boldsymbol{u} \in \mathbb{R}^n$,*

$$\lim_{t \to 0} \frac{f(\boldsymbol{x} + t\boldsymbol{u}) - f(\boldsymbol{x})}{t} = \langle \nabla f(\boldsymbol{x}), \boldsymbol{u} \rangle.$$

The proof follows from the chain rule (§3) applied to the composition $t \mapsto \boldsymbol{x} + t\boldsymbol{u} \mapsto f(\boldsymbol{x} + t\boldsymbol{u})$. The directional derivative measures how fast f changes if we move in direction $\boldsymbol{u}$, and the proposition says it is the inner product of the gradient with $\boldsymbol{u}$.

Corollary 4. *The gradient $\nabla f(\boldsymbol{x})$ points in the direction of steepest increase of f at $\boldsymbol{x}$. The magnitude $\|\nabla f(\boldsymbol{x})\|$ is the maximum rate of increase, attained when $\boldsymbol{u} = \nabla f(\boldsymbol{x}) / \|\nabla f(\boldsymbol{x})\|$.*

Proof. By Cauchy–Schwarz, $\langle \nabla f(\boldsymbol{x}), \boldsymbol{u} \rangle \le \|\nabla f(\boldsymbol{x})\| \|\boldsymbol{u}\| = \|\nabla f(\boldsymbol{x})\|$ for unit $\boldsymbol{u}$, with equality iff $\boldsymbol{u}$ is the normalized gradient. □

The corollary is the geometric backbone of optimization: at any point, the direction of steepest *ascent* of a function is its gradient, and the direction of steepest descent is the negative gradient. The numerical-optimization technique called *gradient ascent* simply iterates "move a small step in the direction of ∇f" until the gradient is small. At a local maximum, the gradient is zero — there is no direction of ascent, by Corollary 4 — which is the first-order necessary condition the next cluster's optimization handout invokes.[1]

Example 5 (Voter's utility gradient)**.** Continuing the spatial-voting example from earlier handouts: a voter with quadratic-loss utility $u_v(\boldsymbol{p}) = -(\boldsymbol{p}-\boldsymbol{x}_v)^\top A(\boldsymbol{p}-\boldsymbol{x}_v)$ has gradient $\nabla u_v(\boldsymbol{p}) = -2A(\boldsymbol{p}-\boldsymbol{x}_v)$. At the voter's ideal point $\boldsymbol{p} = \boldsymbol{x}_v$, the gradient is zero — consistent with $\boldsymbol{x}_v$ being the voter's optimum. At any other policy, the gradient points back toward $\boldsymbol{x}_v$ (modulated by A, which controls the relative weights on different policy dimensions). A candidate moving in the direction of $\nabla u_v(\boldsymbol{p})$ is moving toward the policy that increases the voter's utility fastest from $\boldsymbol{p}$.

3 Total derivatives, Jacobians, the chain rule

The gradient is the right object for a real-valued function of several variables. The natural extension to a vector-valued function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is the *Jacobian* — the matrix of partial derivatives of each output coordinate with respect to each input coordinate. The Jacobian, in turn, gives the right notion of total derivative: a linear approximation to $\boldsymbol{f}$ near a point. The technical content of multivariable differentiation lives in the Jacobian and its properties, especially the *chain rule*, which says how Jacobians compose under function composition.

Definition 6. Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ with component functions $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))$. The *Jacobian* of $\boldsymbol{f}$ at $\boldsymbol{x}$ is the $m \times n$ matrix

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} (\boldsymbol{x}).$$

For $m = 1$, the Jacobian reduces to the row vector $(\nabla f)^\top$.

Definition 7. $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is *(totally) differentiable* at $\boldsymbol{x}$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ with

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{\|\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x}) - L(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0.$$

The linear map L is the *total derivative* of $\boldsymbol{f}$ at $\boldsymbol{x}$, and its matrix in the standard bases is exactly the Jacobian: $L(\boldsymbol{h}) = D\boldsymbol{f}(\boldsymbol{x})\boldsymbol{h}$.

The total derivative is the unique linear map that approximates $\boldsymbol{f}$ near $\boldsymbol{x}$ to first order. It exists when each partial derivative exists and is continuous (the C^1 condition), but a function can have all

[1] The geometric reading of the gradient extends to its perpendicularity to level sets, and the political-economy reading is worth pausing on. A *level set* of f is a set $\{\boldsymbol{x} : f(\boldsymbol{x}) = c\}$ for a fixed value c; for a voter's utility function on a policy space, the level sets are the indifference curves. At any point on a level set, the gradient is perpendicular to the set — the structural reason being that motion along the level set leaves f unchanged, and motion along the gradient changes f as fast as possible. So the gradient of a voter's utility at any policy is perpendicular to her indifference curve through that policy. The implication for spatial-voting analysis is direct: when one wants to characterize policies that are "just preferred" to a given alternative, the gradient gives the direction along which the just-preferred policies clump up. The Plott-style chaos results in two-dimensional spatial models depend on the geometry of indifference-curve gradients, and the formal moves underlying them invoke this perpendicularity at every step. The proposition extends to constrained optimization: at a constrained optimum, the gradient of the objective and the gradient of the binding constraint are parallel (the Lagrange multiplier of the next cluster's static-optimization handout), again from a perpendicularity-of-gradient-to-level-set argument.

partial derivatives at a point without being totally differentiable there — a subtlety we flag and otherwise leave alone. Rudin (1976, Ch. 9) works through the relationship in detail.

The structural payoff of the Jacobian framework is the chain rule.

Theorem 8 (Chain rule)**.** *Let* $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$ *and* $\boldsymbol{f} : \mathbb{R}^p \to \mathbb{R}^m$ *be continuously differentiable, with composition* $\boldsymbol{h} = \boldsymbol{f} \circ \boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$*. Then* $\boldsymbol{h}$ *is continuously differentiable, with Jacobian*

$$D\boldsymbol{h}(\boldsymbol{x}) = D\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \cdot D\boldsymbol{g}(\boldsymbol{x}).$$

The chain rule is the multivariable generalization of $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$. Its content is that the Jacobian of a composition is the matrix product of the Jacobians, taken in the right order: outer-then-inner, evaluated at the appropriate points. The proof (in Rudin (1976, Ch. 9)) uses the linearity of total derivatives and the definition of differentiability.

Example 9 (Chain rule for comparative statics)**.** A candidate's vote share V depends on her policy position $\boldsymbol{p} \in \mathbb{R}^k$, which in turn depends on parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ (her ideological constraints, fundraising network, etc.) via $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{\theta})$. The total derivative of V with respect to $\boldsymbol{\theta}$ is, by the chain rule,

$$DV(\boldsymbol{\theta}) = \nabla V(\boldsymbol{p}(\boldsymbol{\theta}))^\top \cdot D\boldsymbol{p}(\boldsymbol{\theta}),$$

a $1 \times d$ matrix (i.e., a row vector). The ith component, $\frac{\partial V}{\partial \theta_i}$, is the marginal effect of parameter i on vote share, computed as the gradient of V with respect to $\boldsymbol{p}$ dotted with the rate at which $\boldsymbol{p}$ moves in response to θ_i. The decomposition is the formal basis for "the parameter affects the outcome only through its effect on the policy choice" — a structural claim political economists make all the time.

4 Higher-order derivatives and Hessians

Just as the second derivative of a single-variable function captures convexity / concavity and the second-order behavior of the function, the second derivatives of a multivariable function capture its second-order geometry. The right object is the *Hessian* — the matrix of all second partial derivatives. The Hessian is the multivariable analogue of $f''(x)$, and it is what the second-order conditions for optimization, the curvature of utility functions, and the local quadratic approximations all use. The Hessian is also *symmetric* when f is twice continuously differentiable, by the equality of mixed partial derivatives (Clairaut's theorem) — a structural fact that the previous handout's spectral theorem leans on heavily.

Definition 10. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. The *Hessian* of f at $\boldsymbol{x}$ is the $n \times n$ matrix of second partial derivatives

$$H_f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} (\boldsymbol{x}).$$

$f \in C^2$ at $\boldsymbol{x}$ if every second partial is continuous in a neighborhood of $\boldsymbol{x}$.

Theorem 11 (Symmetry of mixed partials, Clairaut)**.** *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is twice continuously differentiable at* $\boldsymbol{x}$*, then for every* i, j*,*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{x}).$$

Equivalently, the Hessian $H_f(\boldsymbol{x})$ *is a symmetric matrix.*

The proof uses the mean value theorem twice, applied to the increments of the partials in opposite orders. Rudin (1976, Ch. 9) works it through. The political-economy upshot: every Hessian one ever computes from a twice-continuously-differentiable utility function is symmetric, so the spectral theorem of the previous handout applies directly — the eigenvalue-based definiteness machinery is exactly the apparatus needed to characterize the function's local curvature.

Example 12 (Hessian of quadratic-loss utility)**.** Continuing Example 5: $u_v(\boldsymbol{p}) = -(\boldsymbol{p} - \boldsymbol{x}_v)^\top A(\boldsymbol{p} - \boldsymbol{x}_v)$ for symmetric A. The Hessian is $H_{u_v}(\boldsymbol{p}) = -2A$, independent of $\boldsymbol{p}$ (the function is exactly quadratic, so the Hessian is constant). The voter's utility is strictly concave iff $-2A$ is negative definite, iff A is positive definite, iff A's eigenvalues are all positive — which (by Sylvester's criterion from the previous handout) is testable on the leading principal minors of A.

5 Taylor's theorem

Taylor's theorem says that a smooth function near a point $\boldsymbol{x}_0$ is well-approximated by a polynomial in the displacement $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{x}_0$, with the polynomial's coefficients given by the function's derivatives at $\boldsymbol{x}_0$. The political-economy use of Taylor's theorem is to localize: when the global function is hard to analyze but the local structure near a point is what matters (an equilibrium, an optimum, a calibration value), the second-order Taylor expansion reduces the analysis to a quadratic form. Most of the local-comparative-statics, second-order-conditions, and curvature arguments in formal political-economy theory are second-order Taylor expansions in disguise.

Theorem 13 (Taylor, second-order in $\mathbb{R}^n$)**.** *Let $f : U \to \mathbb{R}$ be twice continuously differentiable on an open set $U \subseteq \mathbb{R}^n$, and let $\boldsymbol{x}_0 \in U$. For $\boldsymbol{h}$ small enough that $\boldsymbol{x}_0 + \boldsymbol{h} \in U$,*

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^\top \boldsymbol{h} + \tfrac{1}{2}\boldsymbol{h}^\top H_f(\boldsymbol{x}_0)\boldsymbol{h} + r(\boldsymbol{h}),$$

where the remainder $r(\boldsymbol{h})$ satisfies $r(\boldsymbol{h})/\|\boldsymbol{h}\|^2 \to 0$ as $\boldsymbol{h} \to \boldsymbol{0}$.

The proof reduces to the single-variable case along the line $t \mapsto \boldsymbol{x}_0 + t\boldsymbol{h}$ for $t \in [0, 1]$, with the remainder bounded by an integral form of the second-derivative remainder. Rudin (1976, Ch. 9) or Apostol (1974, Ch. 6) works it through.

The structural reading: near $\boldsymbol{x}_0$, f behaves like an affine function (the gradient term) plus a quadratic correction (the Hessian term). When the gradient is zero (i.e., $\boldsymbol{x}_0$ is a critical point), the affine part vanishes, and the second-order behavior of f is governed entirely by the quadratic form $\boldsymbol{h}^\top H_f(\boldsymbol{x}_0)\boldsymbol{h}$ — the same quadratic-form-and-definiteness story the previous handout developed. Sylvester's criterion applied to $H_f(\boldsymbol{x}_0)$ tells us the local maximum / minimum / saddle-point structure of f near a critical point, which is exactly the second-order condition for unconstrained optimization in the next cluster.

6 The implicit function theorem

Most equilibrium objects in formal political-economy theory are defined *implicitly*: a Nash equilibrium is a fixed point of best-response correspondences; a market-clearing price is the solution of a system of supply-equals-demand equations; a Bayesian-game equilibrium is the solution of a system of expected-payoff conditions. In every case, the analyst doesn't have a closed-form expression for the equilibrium as a function of the model's parameters. What she wants, instead, is a structural result that lets her differentiate the equilibrium with respect to a parameter using only the system of equations that defines it. The implicit function theorem is exactly that result, and it is the workhorse of comparative statics in formal theory.

Theorem 14 (Implicit function theorem)**.** *Let $\boldsymbol{F} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be continuously differentiable on a neighborhood of $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, with $\boldsymbol{F}(\boldsymbol{x}_0, \boldsymbol{y}_0) = \boldsymbol{0}$. Suppose the $m \times m$ Jacobian $D_{\boldsymbol{y}}\boldsymbol{F}(\boldsymbol{x}_0, \boldsymbol{y}_0)$ (the matrix of partials of $\boldsymbol{F}$ with respect to the $\boldsymbol{y}$ coordinates only) is invertible. Then there exist neighborhoods U of $\boldsymbol{x}_0$ in $\mathbb{R}^n$ and V of $\boldsymbol{y}_0$ in $\mathbb{R}^m$ and a continuously differentiable function $\boldsymbol{g} : U \to V$ with $\boldsymbol{g}(\boldsymbol{x}_0) = \boldsymbol{y}_0$ such that, for every $\boldsymbol{x} \in U$,*

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x})) = \boldsymbol{0}.$$

Moreover, the Jacobian of $\boldsymbol{g}$ is given by

$$D\boldsymbol{g}(\boldsymbol{x}) = -[D_{\boldsymbol{y}}\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))]^{-1} \cdot D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x})).$$

The reading: the system $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$ defines $\boldsymbol{y}$ implicitly as a function of $\boldsymbol{x}$, locally near $(\boldsymbol{x}_0, \boldsymbol{y}_0)$. The condition for this implicit definition to work is that the Jacobian with respect to $\boldsymbol{y}$ be invertible at the base point. When it is, the comparative-static derivative $D\boldsymbol{g} = D\boldsymbol{y}/D\boldsymbol{x}$ is computable from the formula in the theorem — a matrix expression that uses the system $\boldsymbol{F}$ only, not the (often unavailable) closed form for $\boldsymbol{g}$.

Rudin (1976, Ch. 9) gives the proof via the contraction mapping theorem (which the dynamic-optimization handout will treat in its own right). The structural intuition: linearize $\boldsymbol{F}$ at $(\boldsymbol{x}_0, \boldsymbol{y}_0)$; the linearization is invertible in $\boldsymbol{y}$ by the Jacobian condition, so locally $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$ has a unique solution $\boldsymbol{y}$ for each $\boldsymbol{x}$, and that solution depends differentiably on $\boldsymbol{x}$.[2]

Example 15 (Comparative statics of an equilibrium tax)**.** A government chooses a tax rate τ that satisfies the equilibrium condition $F(\tau, c) = 0$, where c is the cost of public goods. Suppose at (τ_0, c_0) we have $F = 0$ and $\frac{\partial F}{\partial \tau}(\tau_0, c_0) \neq 0$. By the IFT, there exists a function $\tau(c)$ such that $F(\tau(c), c) = 0$ for c near c_0, and

$$\frac{\partial \tau}{\partial c} = -\frac{\frac{\partial F}{\partial c}}{\frac{\partial F}{\partial \tau}}.$$

Without solving for τ explicitly, the analyst can read off the sign of $\partial\tau/\partial c$ from the signs of the two partial derivatives of F. This is the standard pattern of comparative-statics arguments throughout formal-theory political-economy literature.

[2]The implicit function theorem is the structural engine of comparative statics in formal political-economy theory. A typical application: a one-shot game has equilibrium strategies $\boldsymbol{s}^*$ characterized by the FOC system $\boldsymbol{F}(\boldsymbol{s}, \boldsymbol{\theta}) = \nabla_{\boldsymbol{s}} u(\boldsymbol{s}, \boldsymbol{\theta}) = \boldsymbol{0}$, where $\boldsymbol{\theta}$ is a vector of parameters. The IFT, applied at an equilibrium $(\boldsymbol{s}^*, \boldsymbol{\theta}^*)$ where the Hessian $D_{\boldsymbol{s}}\boldsymbol{F} = H_u(\boldsymbol{s}^*, \boldsymbol{\theta}^*)$ is invertible (a generic condition, equivalent to the equilibrium being non-degenerate), gives $D_{\boldsymbol{\theta}}\boldsymbol{s}^* = -[H_u]^{-1} D_{\boldsymbol{\theta}} \nabla_{\boldsymbol{s}} u$. This is the comparative-statics formula one finds in nearly every formal-theory paper, sometimes derived from first principles, sometimes invoked as boilerplate. The signs of the entries of $D_{\boldsymbol{\theta}}\boldsymbol{s}^*$ tell us how each equilibrium component shifts with each parameter, and substantive predictions about the strategic-environment-and-its-shifts are read off from these signs. A structural alternative when the Hessian is positive- or negative-definite (i.e., when the FOC characterizes a unique equilibrium) is the monotone-comparative-statics theory of Milgrom and Shannon (1994) and Topkis (1998), which delivers comparative statics without differentiability via lattice-theoretic arguments. Both approaches lead to the same structural conclusions in the well-behaved cases, with the IFT giving the local-quantitative answer and Topkis-Milgrom-Shannon giving the global-qualitative one.

7 What's next

The next handout takes up convex sets and concave functions. With the gradient and Hessian apparatus of this handout in hand, the structural definitions land naturally: a C^1 function is concave iff its gradient is a supporting linear functional at every point; a C^2 function is concave iff its Hessian is everywhere negative semidefinite (the eigenvalue-and-Sylvester machinery from the previous handout). The handout after that uses both: static optimization writes FOC in gradient form and SOC in Hessian-definiteness form, and the constrained-optimization machinery (Lagrangians, KKT) generalizes the unconstrained gradient condition to settings where the optimum lies on the boundary of a feasible set.

For graduate-level treatments at this handout's level of abstraction, Rudin (1976, Ch. 9) is the canonical reference for multivariable differentiation. Apostol (1974, Ch. 6, 12, 13) treats the same material with somewhat more applied flavor. Spivak (1965) is the classical text on differentiation in the manifold setting (beyond what we need here, but excellent for those who continue further). For the implicit function theorem and its political-economy applications, Sundaram (1996, Ch. 1) works through both the theorem and several PE-style examples; Mas-Colell, Whinston, and Green (1995, App. M) treats the same material with the comparative-statics applications of microeconomic theory.

8 Exercises

Exercise 16. Compute the gradient and Hessian of $f(x_1, x_2) = x_1^2 + 3x_1x_2 + 2x_2^2 - 4x_1 + x_2$. Identify the critical point and determine whether it is a local maximum, minimum, or saddle by computing the eigenvalues of the Hessian.

Exercise 17. At the point $(1, 1)$, in which direction does $f(x_1, x_2) = x_1^2 + x_2^2$ decrease fastest? At what rate? Verify directly using the directional-derivative formula.

Exercise 18. Show that for any continuously differentiable $f : \mathbb{R}^2 \to \mathbb{R}$, the gradient at any point on a level set $\{f = c\}$ is perpendicular to the tangent to the level set at that point. (Hint: parametrize the level set near the point and differentiate.) Interpret in spatial-voting terms: at any policy on a voter's indifference curve, the voter's gradient is perpendicular to the indifference curve.

Exercise 19. Let $\boldsymbol{g}(t) = (t^2, t^3)$ and $f(x_1, x_2) = x_1 + x_2^2$. Compute $\frac{d}{dt} f(\boldsymbol{g}(t))$ in two ways: (a) by composing and differentiating directly; (b) via the chain rule, $\frac{\partial f}{\partial x_1} g_1'(t) + \frac{\partial f}{\partial x_2} g_2'(t)$. Verify the two methods agree.

Exercise 20. *Best-response Jacobian in a two-player game.* Each player's best response is given by an FOC. Suppose the system is $F_1(s_1, s_2) = 0$, $F_2(s_1, s_2) = 0$ at an equilibrium (s_1^*, s_2^*). Write the Jacobian $D\boldsymbol{F}$ as a 2×2 matrix in terms of partial derivatives. Under what condition is the equilibrium "locally unique" (i.e., the only one in a neighborhood)?

Exercise 21. Compute the second-order Taylor expansion of $f(x_1, x_2) = e^{x_1} \cos x_2$ at $(0, 0)$. Verify that the gradient and Hessian terms agree with direct computation from Definition 2 and Definition 10.

Exercise 22. Verify Clairaut's theorem on a small example. Compute the mixed partials $\partial^2 f / \partial x_1 \partial x_2$ and $\partial^2 f / \partial x_2 \partial x_1$ for $f(x_1, x_2) = x_1^3 x_2^2 + x_1 \sin x_2$, and confirm they are equal.

Exercise 23. *Implicit function theorem, direct application.* The equation $x^2 + y^2 = 1$ defines y implicitly as a function of x near $(0, 1)$. (a) Verify the conditions of the IFT at this point. (b) Compute dy/dx at this point using the IFT formula. (c) Verify by solving for y explicitly and differentiating.

Exercise 24. *Comparative statics via IFT.* An equilibrium is defined by the system $F_1(s_1, s_2, \theta) = s_1 - s_2^2 - \theta = 0$, $F_2(s_1, s_2, \theta) = s_1 + s_2 + \theta^2 - 2 = 0$. (a) At $\theta = 0$, find the equilibrium (s_1^*, s_2^*). (b) Use the IFT to compute $\partial s_1^*/\partial\theta$ and $\partial s_2^*/\partial\theta$ at this equilibrium. (c) Discuss in one sentence why the IFT is the right tool here, rather than just solving the system explicitly for general θ.

Exercise 25. *Vote-share comparative statics.* A candidate's vote share $V(\boldsymbol{p}, \boldsymbol{\theta}) = -(\boldsymbol{p} - \boldsymbol{\theta})^\top(\boldsymbol{p} - \boldsymbol{\theta})$ depends on her position $\boldsymbol{p} \in \mathbb{R}^2$ and a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^2$ (interpretable as the median voter's ideal point). Suppose she chooses $\boldsymbol{p}$ to maximize V for given $\boldsymbol{\theta}$. (a) Compute $\boldsymbol{p}^*(\boldsymbol{\theta})$ via the FOC. (b) Compute $D\boldsymbol{p}^*(\boldsymbol{\theta})$ via the IFT applied to the FOC system. (c) Verify that the IFT result agrees with direct differentiation of $\boldsymbol{p}^*$.

References

Apostol, Tom M. (1974). *Mathematical Analysis.* 2nd ed. Reading, MA: Addison-Wesley.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995). *Microeconomic Theory.* New York: Oxford University Press.

Milgrom, Paul and Chris Shannon (1994). "Monotone Comparative Statics". In: *Econometrica* 62.1, pp. 157–180.

Rudin, Walter (1976). *Principles of Mathematical Analysis.* 3rd ed. New York: McGraw-Hill.

Spivak, Michael (1965). *Calculus on Manifolds.* New York: W. A. Benjamin.

Sundaram, Rangarajan K. (1996). *A First Course in Optimization Theory.* Cambridge: Cambridge University Press.

Topkis, Donald M. (1998). *Supermodularity and Complementarity.* Princeton: Princeton University Press.