ARE201-Simon, Fall2015

CALCULUS3: TUE, SEP 15, 2015 PRINTED: SEPTEMBER 15, 2015 (LEC# 6)

Contents

2. Univariate and Multivariate Differentiation (cont)

2.4. Multivariate Calculus: functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

We'll now generalize what we did last time to a function $f : \mathbb{R}^n \to \mathbb{R}^m$. In general, if you have a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, what is the notion of slope (or gradient or derivative)? Not suprisingly, it is a $m \times n$ *matrix*. The *matrix* which is the derivative of a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is called the *Jacobian matrix* for that function.

Note well: When I talk about the Jacobian of a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, I'm referring to the matrix which is the function's derivative. When $n = m$, the Jacobian has a *determinant*, properly called the Jacobian determinant. However, there are some books that use the unqualified term Jacobian to refer to the Jacobian determinant. So you need to be aware of which is which.

Our three stages of calculus:

(1) Calculus: first stage: If $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then we have
   (a) The derivative of a function at a point, $f'(x) \in \mathbb{R}$.
   (b) The derivative $f' : \mathbb{R} \to \mathbb{R}$ is a (generally) nonlinear function
   (c) The differential function, $Lf^x(dx) = f'(x)dx$, i.e., $Lf^x : \mathbb{R} \to \mathbb{R}$ is a linear function
   This illustrates the isomorphism between $L(\mathbb{R}, \mathbb{R})$and $\mathbb{R}$, i.e.,

   $$Lf^x = f'(x)dx \quad : \quad f'(x) \quad :: \quad y = \alpha x \quad : \quad \alpha \in \mathbb{R}$$

(2) Calculus: second stage: if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then we have
   (a) The derivative of a function at a point, $\bigtriangledown f(\mathbf{x}) \in \mathbb{R}^n$.
   (b) The derivative $\bigtriangledown f : \mathbb{R}^n \to \mathbb{R}^n$ is a (generally) nonlinear function
   (c) The differential function, $Lf^{\mathbf{x}}(\mathbf{dx}) = \bigtriangledown f(\mathbf{x})\mathbf{dx}$, i.e., $Lf^{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}$ is a linear function
   This illustrates the isomorphism between $L(\mathbb{R}^n, \mathbb{R})$and $\mathbb{R}^n$, i.e.,

   $$Lf^{\mathbf{x}} = \bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx} \quad : \quad \bigtriangledown f(\mathbf{x}) \quad :: \quad y = \boldsymbol{\alpha} \cdot \mathbf{x} \quad : \quad \boldsymbol{\alpha} \in \mathbb{R}^n$$

(3) Calculus: third stage: If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, then we have
   (a) The derivative of a function at a point, $\mathrm{Jf}(\mathbf{x}) \in \mathbb{R}^{m \times n}$.
   (b) The derivative $\mathrm{Jf} : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ is a (generally) non-linear function
   (c) The differential function, $\mathbf{Lf}^{\mathbf{x}}(\mathbf{dx}) = \mathrm{Jf}(\mathbf{x})\mathbf{dx}$, i.e., $\mathbf{Lf} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function
   This illustrates the isomorphism between $L(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathbb{R}^{m \times n}$, i.e.,

   $$Lf^{\mathbf{x}} = \mathrm{Jf}(\mathbf{x})\mathbf{dx} \quad : \quad \mathrm{Jf}(\mathbf{x}) \quad :: \quad \mathbf{y} = M\mathbf{x} \quad : \quad M$$

Example: A particularly important function from $\mathbb{R}^n$ to $\mathbb{R}^n$ is the gradient of $f : \mathbb{R}^n \to \mathbb{R}$. Specifically, think of the gradient as being n functions from $\mathbb{R}^n \to \mathbb{R}$, i.e., each of the partial derivatives of f, stacked on top of each other: $\bigtriangledown f = \begin{bmatrix} f_1(\cdot) \\ \vdots \\ f_n(\cdot) \end{bmatrix}$. The derivative of the gradient function is the matrix constructed by stacking the gradients of each of these partial derivatives *viewed as row vectors* on top of each other, i.e., $\begin{bmatrix} \bigtriangledown f_1(\cdot) \\ \vdots \\ \bigtriangledown f_n(\cdot) \end{bmatrix}$. This derivative of the derivative of $f : \mathbb{R}^n \to \mathbb{R}$, which in generic language would be called the Jacobian of the gradient of f, is more concisely known as the *Hessian* of f.

More generally, to visualize the derivative and differential associated with an *arbitrary* function $f : \mathbb{R}^n \to \mathbb{R}^m$, it is helpful to think of f, once again, as a vertical stack of m functions $f^i : \mathbb{R}^n \to \mathbb{R}$, all stacked on top of each other. (Notationaly, the only difference between this and the previous paragraph is that now we use superscripts rather than subscripts to distinguish the functions from each other.) It is now natural to think of the derivative of f as a vertical stack of all the derivatives (gradients) of the f^i's. That is, $f'(\cdot) \equiv \mathrm{Jf}(\cdot) = \begin{bmatrix} \bigtriangledown f^1(\cdot) \\ \bigtriangledown f^2(\cdot) \\ \vdots \\ \bigtriangledown f^m(\cdot) \end{bmatrix}$, where each $\bigtriangledown f^i(\cdot)$ is a row vector consisting of the partial derivatives of $f^i(\cdot)$.

Next think of the *differential* of $\bigtriangledown f$ at $\mathbf{x}$, i.e., the linear function $\mathbf{Lf^x}(\cdot) = \text{Jf}(\mathbf{x})(\cdot)$ as a vertical stack consisting of the differentials of the f^i's at $\mathbf{x}$, i.e.,

$$\mathbf{Lf^x}(\mathbf{dx}) \quad = \quad \text{Jf}(\mathbf{x})(\mathbf{dx}) \quad = \quad \text{Jf}(\mathbf{x}) \cdot \mathbf{dx} \quad = \quad \begin{bmatrix} \bigtriangledown f^1(\mathbf{x}) \cdot \mathbf{dx} \\ \bigtriangledown f^2(\mathbf{x}) \cdot \mathbf{dx} \\ \vdots \\ \bigtriangledown f^m(\mathbf{x}) \cdot \mathbf{dx} \end{bmatrix}.$$

2.5. **Four graphical examples.**

We can now apply all the graphical intuitions we've developed from the last lecture about the differential of a real-valued function, to the general case: instead of considering one 3-D picture like Figure 1 in the previous lecture, you just visualize a stack of m such pictures.

The following example is intended to illustrate this idea. We start out with a function $f : \mathbb{R}^2 \to \mathbb{R}$. Its gradient, then, maps $\mathbb{R}^2$ to $\mathbb{R}^2$. The function we are interested in is graphed in Fig. 1. Note that the function *decreases* with both arguments so that the gradient is a strictly negative vector. We are interested in how the gradient changes in response to a small change $\mathbf{dx}$ in the domain.

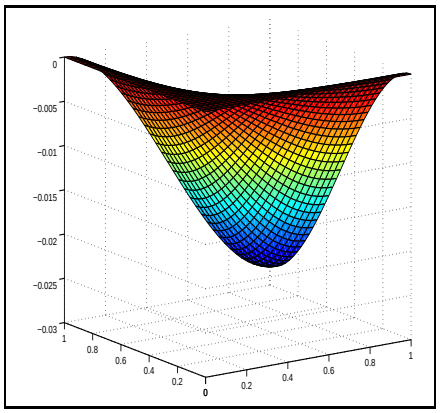

FIGURE 1. Graph of f

To get some intuition, it's helpful to return to the 3-D diagrams that we were looking at in the last lecture, as we do in Fig. 2 below.

The function being graphed in Fig. 1 is

$$f(\mathbf{x}) \quad = \quad \big(x_1^2/2 - x_1^3/3\big)\big(x_2^3/3 - x_2^2/2\big)$$

whose gradient is

$$\bigtriangledown f(\mathbf{x}) \quad = \quad \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{bmatrix} \quad = \quad \begin{bmatrix} \big(x_1 - x_1^2\big)\big(x_2^3/3 - x_2^2/2\big) \\ \big(x_2^2 - x_2\big)\big(x_1^2/2 - x_1^3/3\big) \end{bmatrix}$$

so that $d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) = \mathbf{Hf}(\mathbf{x}) \cdot \mathbf{dx} = \begin{bmatrix} \bigtriangledown f_1(\mathbf{x}) \\ \bigtriangledown f_2(\mathbf{x}) \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}$, where

$$\bigtriangledown f_1(\mathbf{x}) \quad = \quad \big[\big(1 - 2x_1\big)\big(x_2^3/3 - x_2^2/2\big) \qquad \big(x_1 - x_1^2\big)\big(x_2^2 - x_2\big)\big] \qquad \text{and}$$

$$\bigtriangledown f_2(\mathbf{x}) \quad = \quad \big[\big(x_2^2 - x_2\big)\big(x_1 - x_1^2\big) \qquad \big(2x_2 - 1\big)\big(x_1^2/2 - x_1^3/3\big)\big]$$

We'll evaluate the gradient of this function at the point $\mathbf{x} = [0.667,\ 0.667]$, and consider a shift in the domain of $\mathbf{dx} = [-0.1944,\ 0.2222]$, which takes us to the point $\mathbf{x} + \mathbf{dx} = [0.4722,\ 0.8889]$.

Plugging in the numbers, we obtain

$$\bigtriangledown f(\mathbf{x}) = \begin{bmatrix} -0.0274 \\ -0.0274 \end{bmatrix}; \quad \bigtriangledown f(\mathbf{x}+\mathbf{dx}) = \begin{bmatrix} -0.0401 \\ -0.0075 \end{bmatrix} \quad \text{so that} \quad \bigtriangledown f(\mathbf{x}+\mathbf{dx}) - \bigtriangledown f(\mathbf{x}) = \begin{bmatrix} -0.0127 \\ 0.0199 \end{bmatrix}$$

i.e., the first partial becomes *more* negative while the second becomes *less* so. Evaluating the differential of $\bigtriangledown f$ at $\mathbf{x}$ at the magnitude of the change we obtain

$$d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) \quad = \quad \mathbf{Hf}(\mathbf{x}) \cdot \mathbf{dx} \quad = \quad \begin{bmatrix} 0.0412 & -0.0494 \\ -0.0494 & 0.0412 \end{bmatrix} \begin{bmatrix} -0.1944 \\ 0.2222 \end{bmatrix} \quad = \quad \begin{bmatrix} -0.0190 \\ 0.0187 \end{bmatrix}$$

Note that when we evaluate the differential, the second component of the approximation is much closer to the second component of the true change in $\bigtriangledown f$ than is the first element.

To see the graphical analog of these computations, we'll now do exactly what we were doing for a function mapping $\mathbb{R}^2$ to $\mathbb{R}$, except that we are going to look at two 3-D graphs simultaneously. **It's much easier to understand Fig. 2 if can view it in color, so if you don't have access to a color printer, you might want to look at it on a color screen.** Here's a guide to the colors:

- The *level* of $\bigtriangledown f(\mathbf{x})$ is indicated by pink lines;
- The *level* of $\bigtriangledown f(\mathbf{x}+\mathbf{dx})$ is indicated by purple lines
- The *true change* in $\bigtriangledown f(\cdot)$ is indicated by green lines;
- The *evaluation of the differential* is indicated by red lines

Observe in Fig. 2 that because of the shape of $f_2(\cdot)$, the first order linear approximation to $f_2(\mathbf{x}+\mathbf{dx})$ is almost perfect, while the first order linear approximation to $f_1(\mathbf{x}+\mathbf{dx})$ is much less so. This is reflected in the bottom right panel, where there is a big gap between $(f_1(\mathbf{x}+\mathbf{dx}) - f_1(\mathbf{x}))$ and $df_1^{\mathbf{x}}(\mathbf{dx})$ and a negligible one between $(f_2(\mathbf{x}+\mathbf{dx}) - f_2(\mathbf{x}))$ and $df_2^{\mathbf{x}}(\mathbf{dx})$.

We now consider three more examples, using the differential of the gradient of f to explore how the gradient vector changes as we change $\mathbf{x}$. Since the gradient of f at $\mathbf{x}$ is always perpendicular to the level set of f corresponding to $f(\mathbf{x})$, what we learn about these changes indirectly tells us about things like the curvature of the level set of f at $\mathbf{x}$. Here are a couple of examples, applied to the function $f(\mathbf{x}) = x_1 x_2$.

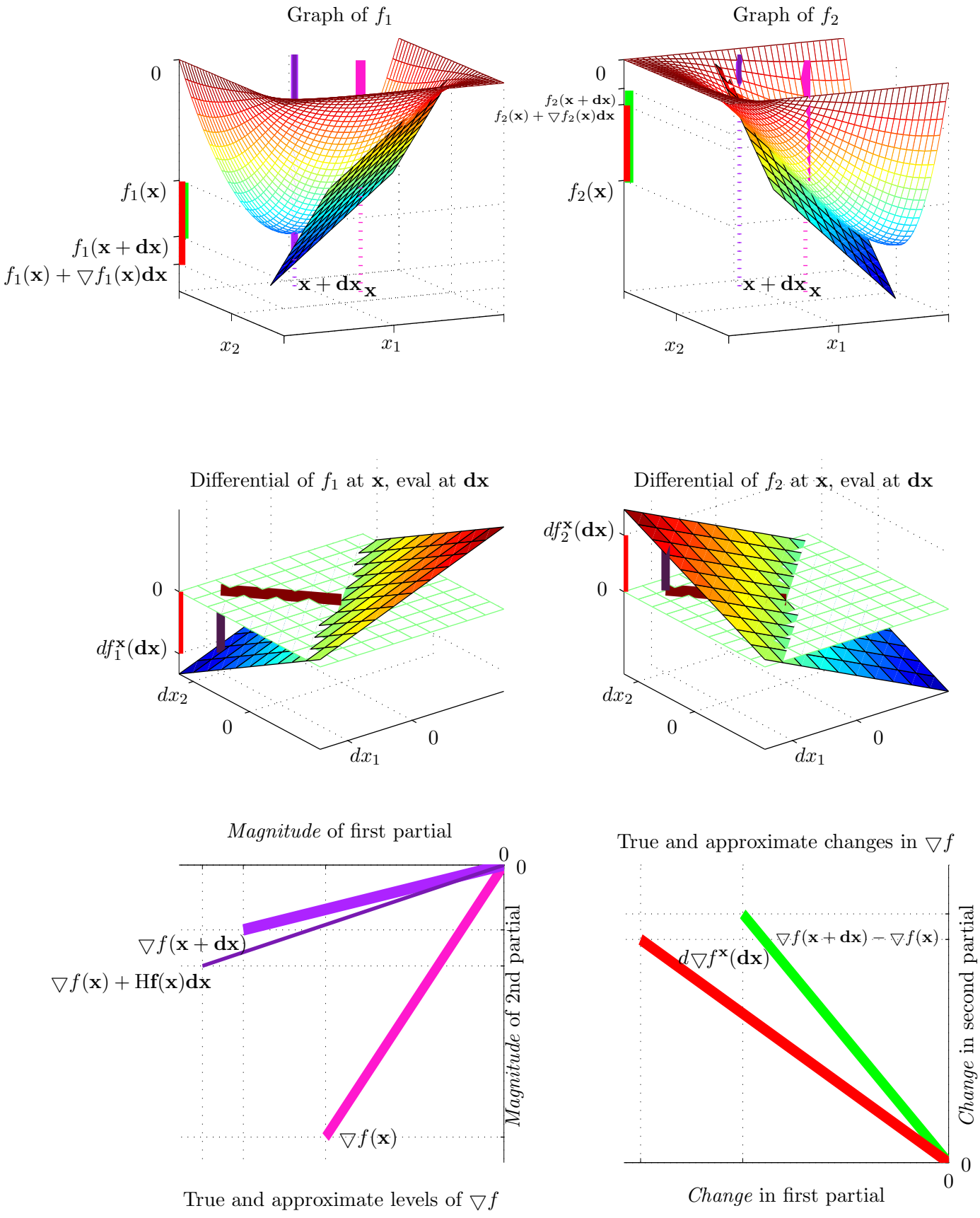

FIGURE 2. The differential approximation to a change in gradient

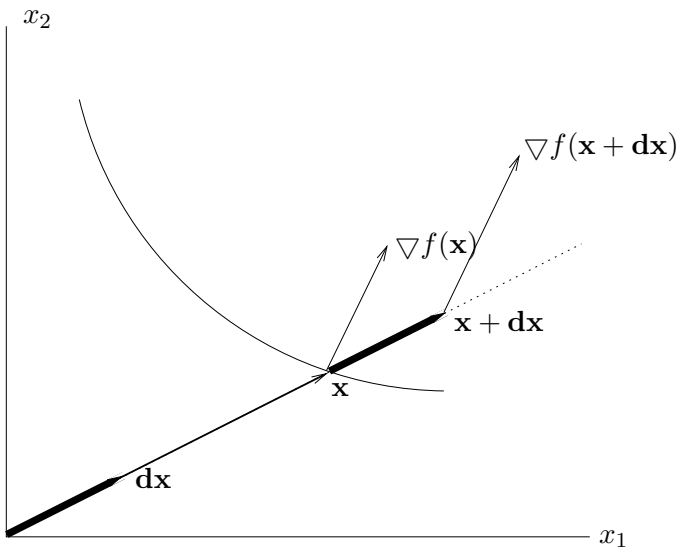

FIGURE 3. f is homothetic

Second example: The function $f(\mathbf{x}) = x_1x_2$, depicted in Fig. 3, is an example of a *homothetic* function, i.e., a function with the property that the *slopes* of its level sets are constant along rays through the origin. More precisely, if $y = \alpha x$, for some scalar $\alpha \in \mathbb{R}_+$, then the slope of the level set of f through $\mathbf{y}$ is equal to the slope of the level set of f through $\mathbf{x}$. Since gradient vectors are perpendicular to level sets, this implies that the gradients of f at both $\mathbf{x}$ and $\mathbf{y}$ must point in the same direction. Let's check that this is true for this function.

$$\begin{aligned} \bigtriangledown f(\mathbf{x}) &= \begin{bmatrix} x_2 & x_1 \end{bmatrix} \\ \mathbf{Hf}(\mathbf{x}) = J \bigtriangledown f(\mathbf{x}) &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned}$$

so the differential of $\bigtriangledown f$ at $\mathbf{x}$ is

$$d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}$$

In this case $J \bigtriangledown f(\mathbf{x})$ is a constant, so that the higher order terms in the Taylor approx are all zero, so that the first approximation must be exactly correct. Now consider a move $\mathbf{dx}$ along the ray through the origin passing through $\mathbf{x}$, i.e., choose $\mathbf{dx} = \alpha\mathbf{x}$, for some scalar $\alpha > 0$. In this case, we have

$$d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix} = \begin{bmatrix} \alpha x_2 \\ \alpha x_1 \end{bmatrix}$$

so that, taking a first order approximation to $\bigtriangledown f(\mathbf{x} + \alpha\mathbf{x})$:

$$\bigtriangledown f(\mathbf{x} + \alpha\mathbf{x}) \approx \bigtriangledown f(\mathbf{x}) + d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) = \begin{bmatrix} (1+\alpha)x_2 \\ (1+\alpha)x_1 \end{bmatrix}$$

But in this case, we can replace the approximation symbol with an equality. That is, the gradient of f at $(1 + \alpha)\mathbf{x}$ is a scalar multiple of the gradient of f at $\mathbf{x}$, confirming homotheticity. (Note additionally that the gradient gets *longer* as you move out along a ray through the origin, indicating that f exhibits increasing returns to scale.)

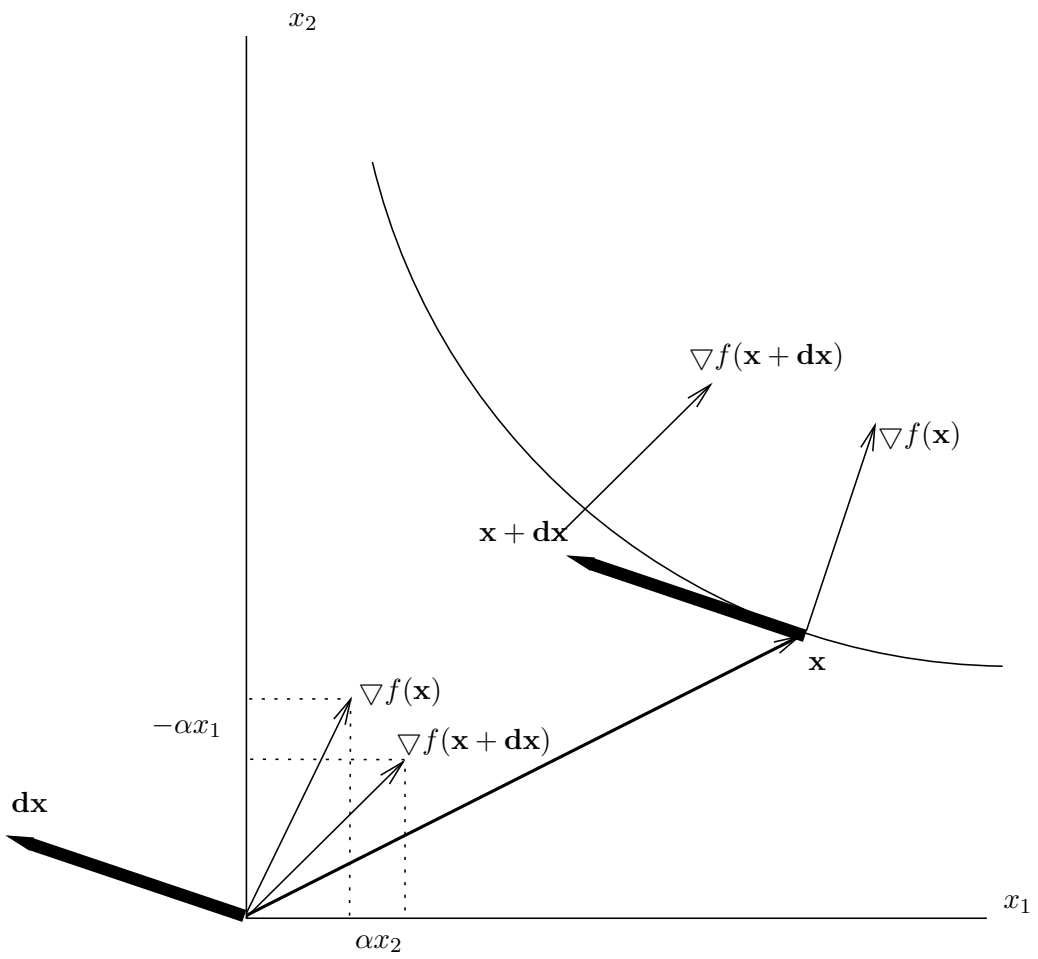

FIGURE 4. f exhibits diminishing MRS

Third example (see Fig. 4): We'll now show that $f(x) = x_1 x_2$ exhibits diminishing marginal rate of substitution. Recall that the marginal rate of substitution of x_2 for x_1 is the ratio $\left|\frac{f_1(x)}{f_2(x)}\right|$. In Fig. 4, this is the length of the horizontal component of the gradient vector divided by the length of the vertical component. i.e., "run over rise." Diminishing MRS means that the gradient vector becomes flatter (steeper) as we move to the northwest (south east) along a level set. We consider a northwesterly movement of x, and verify that the gradient vector becomes flatter. Fix $\mathbf{x}$ and consider a north-west movement in the domain, orthogonal to the gradient of f. Since, $\bigtriangledown f(\mathbf{x})' = [x_2, x_1]$, a north-west movement orthogonal to this vector would be $\mathbf{dx} = \begin{bmatrix} -\alpha x_1 \\ \alpha x_2 \end{bmatrix}$. (Observe that $\bigtriangledown f(\mathbf{x}) \cdot \mathbf{dx} = -\alpha x_1 x_2 + \alpha x_1 x_2 = 0$, so that indeed $\mathbf{dx}$ and $\bigtriangledown f(\mathbf{x})$ are orthogonal to each other). Now

$$d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) \quad = \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -\alpha x_1 \\ \alpha x_2 \end{bmatrix} \quad = \quad \begin{bmatrix} \alpha x_2 \\ -\alpha x_1 \end{bmatrix}$$

so that, evaluating the differential at $\mathbf{dx}$

$$\bigtriangledown f(\mathbf{x}+\alpha\mathbf{x}) \quad \approx \quad \bigtriangledown f(\mathbf{x}) + d\bigtriangledown f^{\mathbf{x}}(\mathbf{dx}) \quad = \quad \begin{bmatrix} (1+\alpha)x_2 \\ (1-\alpha)x_1 \end{bmatrix}$$

or, in other words

$$\bigtriangledown f(\mathbf{x}+\alpha\mathbf{x}) \quad \approx \quad \bigtriangledown f(\mathbf{x}) + \begin{bmatrix} \alpha x_2, \\ -\alpha x_1 \end{bmatrix} \qquad \text{(see fig. Fig. 4)}$$

i.e., the partial with respect to x_1 gets bigger while the partial with respect to x_2 gets smaller, i.e., the gradient gets *flatter*. I've used the approximation symbol above, but as in the preceding example, the differential in this case gives *exactly* the right answer, not just an approximation. As we shall see when we do Taylor theory, the reason we get exactly the right answer in both cases is that the Hessian is independent of $\mathbf{x}$. *However,* it gives exactly the right answer to the gradient at a location that's *not* the one we're interested in. It tells us what the gradient is at $\mathbf{x}+\alpha\mathbf{x}$, which is close to, but not equal to the point on the level set that we're interested in—the point $((1-\alpha)x_1, x_2')$ that lies on the appropriate level set of f—which is vertically above $\mathbf{x}+\alpha\mathbf{x}$.

Fourth example: Consider the demand system: $\boldsymbol{\xi}(\mathbf{p}) = \begin{pmatrix} \xi^1(\mathbf{p}) \\ \vdots \\ \xi^n(\mathbf{p}) \end{pmatrix}$. The Jacobian of this function is written as $J\boldsymbol{\xi}(\cdot)$. *Note that I'm using superscripts rather than subscripts, to distinguish between the components of an* arbitrary *vector-valued function (here the system of demand equations) and the* specific *vector valued function which is the gradient, i.e., vector of partial derivatives.* Start out at $\bar{\mathbf{p}}$. Want to know the effect of a change in the price vector from $\bar{\mathbf{p}}$ to $\mathbf{p}$:

$$\begin{aligned} & \boldsymbol{\xi}(\mathbf{p}) - \boldsymbol{\xi}(\bar{\mathbf{p}}) \\ \approx \; & d\boldsymbol{\xi} \\ = \; & J\boldsymbol{\xi}(\bar{\mathbf{p}})(\mathbf{p}-\bar{\mathbf{p}}) \end{aligned}$$

Explain that $J\boldsymbol{\xi}(\cdot)$ is the matrix constructed by stacking on top of each other the gradients of each of the demand functions. i.e.,

$$J\boldsymbol{\xi}(\bar{\mathbf{p}}) \;=\; \begin{bmatrix} \bigtriangledown\xi^1(\bar{\mathbf{p}})' \\ \vdots \\ \bigtriangledown\xi^n(\bar{\mathbf{p}})' \end{bmatrix}$$

To do a specific example, we are going to set $n = m = 2$. Start out with a given vector $\bar{\mathbf{p}}$, then move it to $\mathbf{p}$. We are interested in approximating the *difference* between the values of the nonlinear function $\boldsymbol{\xi}$, evaluated at these two vectors, i.e., $\boldsymbol{\xi}(\mathbf{p}) - \boldsymbol{\xi}(\bar{\mathbf{p}}) = (dp_1, dp_2)$. We have

$$\begin{aligned} d\boldsymbol{\xi} &= \begin{pmatrix} d\xi^1 \\ d\xi^2 \end{pmatrix} \\ &= \begin{bmatrix} \bigtriangledown\xi^1(\bar{\mathbf{p}})' \\ \bigtriangledown\xi^2(\bar{\mathbf{p}})' \end{bmatrix} \begin{pmatrix} dp_1 \\ dp_2 \end{pmatrix} \\ &= \begin{bmatrix} \xi_1^1(\bar{\mathbf{p}}) & \xi_2^1(\bar{\mathbf{p}}) \\ \xi_1^2(\bar{\mathbf{p}}) & \xi_2^2(\bar{\mathbf{p}}) \end{bmatrix} \begin{pmatrix} dp_1 \\ dp_2 \end{pmatrix} \\ &= \begin{pmatrix} \xi_1^1(\bar{\mathbf{p}})dp_1 + \xi_2^1(\bar{\mathbf{p}})dp_2 \\ \xi_1^2(\bar{\mathbf{p}})dp_1 + \xi_2^2(\bar{\mathbf{p}})dp_2 \end{pmatrix} \end{aligned}$$

Emphasize again that what's going on in all of these examples is that we are approximating the true effect of a change in some variable by the value of the differential, evaluated at the change, in this case a vector.

Do a concrete example with real numbers.

$$\begin{aligned} \boldsymbol{\xi}(\mathbf{p}) &= \begin{pmatrix} y/2p_1 \\ y/2p_2 \end{pmatrix} \\ J\boldsymbol{\xi}(\cdot) &= \begin{bmatrix} -y/2p_1^2 & 0 \\ 0 & -y/2p_2^2 \end{bmatrix} \end{aligned}$$

Set $y = 8000$; $\bar{p}_1 = \bar{p}_2 = 4$; $p_1 = p_2 = 4.1$, so that $\boldsymbol{\xi}(\bar{\mathbf{p}}) = (1000, 1000)$; $\boldsymbol{\xi}(\mathbf{p}) = (975.6, 975.6)$;

Thus $\mathbf{p} - \bar{\mathbf{p}} = (0.1, 0.1)$ while $\boldsymbol{\xi}(\mathbf{p}) - \boldsymbol{\xi}(\bar{\mathbf{p}}) = (-24.4, -24.4)$.

Calculate the approximation:

$$\begin{aligned} d\boldsymbol{\xi}(\cdot) &= \begin{bmatrix} -y/2p_1^2 & 0 \\ 0 & -y/2p_2^2 \end{bmatrix} \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} \\ &= \begin{bmatrix} -8000/32 & 0 \\ 0 & -8000/32 \end{bmatrix} \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} \\ &= \begin{bmatrix} -250 & 0 \\ 0 & -250 \end{bmatrix} \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} \\ &= (-25, -25) \end{aligned}$$

So the approximation is within about 2.5% of the right answer.

Graphically, what is going on here is very similar to what we did in the linear algebra section. That is, we are going to look at the image of $\mathbf{dp}$ under the linear function defined by the Jacobian matrix. Fig. 5 shows the change in price in $\mathbf{dp}$ space, the pair of gradient vectors, the image of $\mathbf{dp}$ under the linear function defined by the Jacobian matrix, and finally the original demand vector together with the approximate location of the new demand vector.

- top left picture is a circle of dp's. The horizontal axis is the first component of $\mathbf{dp}$, the vertical axis is the second.
- bottom left picture has the columns of the Jacobian: emphasize that the vectors are each gradient vectors for each demand function. The label on the horizontal axis is: derivative w.r.t. first price.
- bottom right is what happens to the dp's under the function defined by the Jacobian matrix.
- Top right is in quantity space, and show that where the quantity ends up is roughly obtained by adding the image of dp to the starting vector.

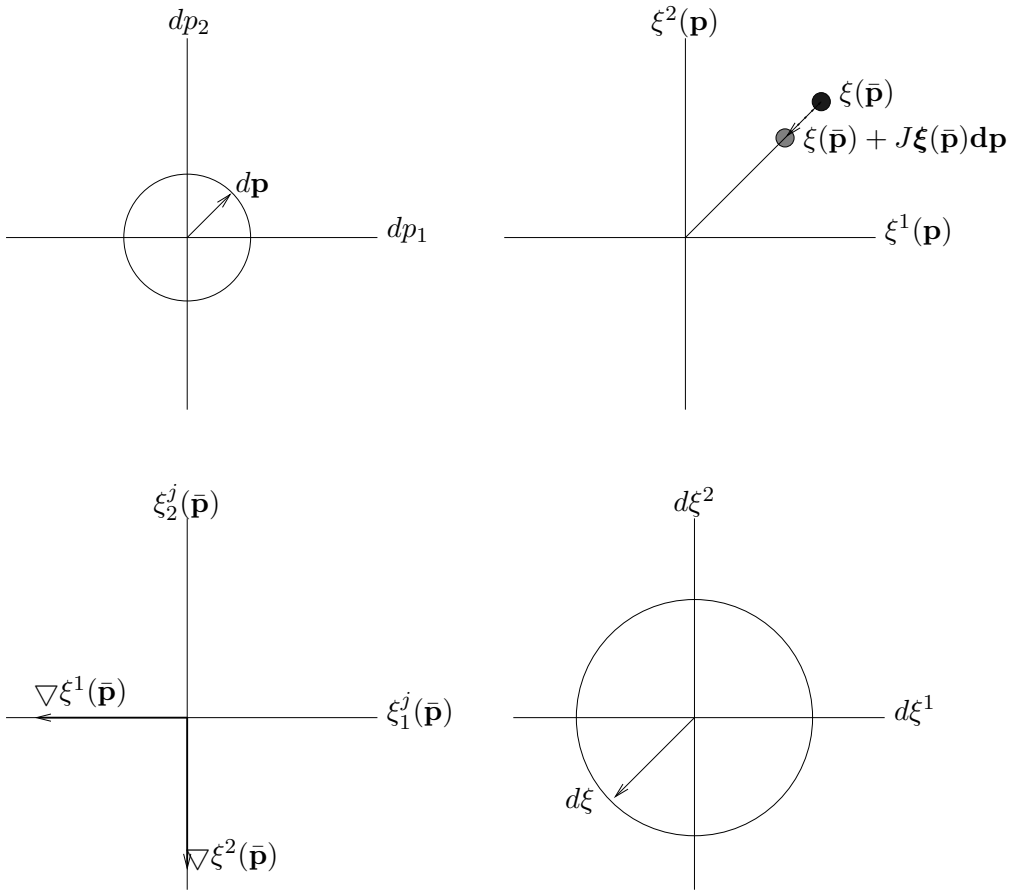

FIGURE 5. Demand as a function of price

2.6. **Taylor's Theorem**

Approximating the change in a nonlinear function by evaluating the differential is only a good approximation if the change is small. As we noted last time, we can improve our approximation by adding in extra terms; instead of doing a *linear* or *first-order* approximation, can do a *quadratic* or *second-order* approximation. Consider Fig. 6. The function $f(\cdot)$ is quite well approximated by the affine function $A(\cdot)$, but it is better approximated by the quadratic function $Q(\cdot)$. And would be better still approximated by a cubic function, etc.

$$f(\bar{x} + dx) - f(\bar{x}) \qquad \approx \qquad f'(\bar{x})dx \tag{1a}$$

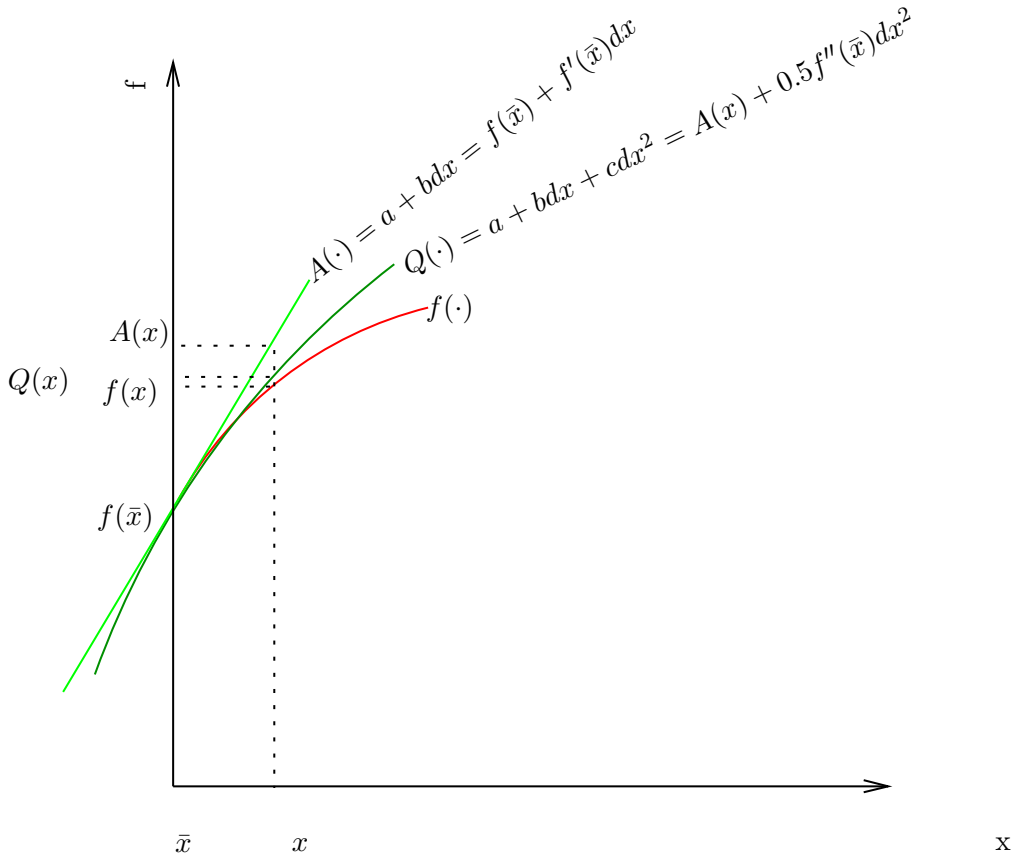

FIGURE 6. Approximating a function with linear and quadratic functions

$$f(\bar{x}+dx) - f(\bar{x}) \underset{\text{even better}}{\approx} f'(\bar{x})dx + \frac{1}{2}f''(\bar{x})dx^2 \tag{1b}$$

$$f(\bar{x}+dx) - f(\bar{x}) = f'(\bar{x})dx + \frac{1}{2}f''(\bar{x})dx^2 + \text{ a remainder term} \tag{1c}$$

We don't need to stop at f'', we can go on forever, the k'th term in the series will be $f^{(k)}(\bar{x})dx^k/k!$, where $f^{(k)}$ denotes the k'th derivative of f (e.g., $f^{(3)} = f'''$) and $k!$ denotes "n-factorial," i.e., $k! = k \times (k-1) \times (k-2) \times ... \times 2$. Note that the equality in (1c) is true trivially. What makes (1c) useful is that we can say something quite specific about the functional form of the "remainder term, " as we'll see, it's the $(k+1)$'th order derivative of f, evaluated at a point *somewhere between* $\bar{x}$ and $\bar{x}+dx$, multiplied by dx^{k+1}.

Similarly, if $f:\mathbb{R}^n \to \mathbb{R}$ and is twice continuously differentiable, then

$$f(\bar{\mathbf{x}}+\mathbf{dx}) - f(\bar{\mathbf{x}}) \quad \approx \quad \bigtriangledown f(\bar{\mathbf{x}})\cdot \mathbf{dx} \tag{2a}$$

$$f(\bar{\mathbf{x}}+\mathbf{dx}) - f(\bar{\mathbf{x}}) \quad \underset{\text{even better}}{\approx} \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} \tag{2b}$$

$$f(\bar{\mathbf{x}}+\mathbf{dx}) - f(\bar{\mathbf{x}}) \quad = \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \frac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} + \text{ a remainder term} \tag{2c}$$

We'll refer to (2c) as a second order Taylor expansion of f about $\bar{\mathbf{x}}$ in the direction $\mathbf{dx}$. The relationship betweeen the left-hand and right-hand sides of (2c) is the content of what's known as Taylor Theory. We'll study two theorems which make precise the notion that the right-hand is a useful way of reformulating the left-hand side.

(1) The first theorem—I'll call it "global Taylor"—specifies the functional form of the remainder term
(2) The second theorem—I'll call it "local Taylor"—identifies conditions under which we can, in a very special sense, "ignore" the remainder term series.

The "global" version of Taylor's theorem is known as the Taylor-Lagrange theorem. I'm going to state it generally for functions mapping $\mathbb{R}$ to $\mathbb{R}$ and as two special cases for functions mapping $\mathbb{R}^n$ to $\mathbb{R}$. The reason for this difference is:

- when $f:\mathbb{R}\to\mathbb{R}$, then all of f's derivatives are scalars.
- when $f:\mathbb{R}^n\to\mathbb{R}$, then the first derivative is a vector, the second a matrix, the third is a hyper-matrix, so you need different notations to deal with first-order Taylor expansions, second-order expansions, etc., etc.

Theorem (Taylor-Lagrange or Global Taylor): If $f:\mathbb{R}\to\mathbb{R}$ is $(K+1)$ times continuously differentiable, then for any $0\leq k\leq K$ and any $\bar{x}, dx\in\mathbb{R}$, there exists $\lambda\in[0,1]$ such that

$$f(\bar{x}+dx) - f(\bar{x}) \quad = \quad f'(\bar{x})dx + \frac{f''(\bar{x})dx^2}{2} + ... + \frac{f^{(k)}(\bar{x})dx^k}{k!} \quad + \quad \frac{f^{(k+1)}(\bar{x}+\lambda dx)dx^{k+1}}{(k+1)!} \tag{3}$$

If $f:\mathbb{R}^n\to\mathbb{R}$ is twice continuously differentiable, then $\forall \bar{\mathbf{x}}, \mathbf{dx}\in\mathbb{R}^n$, $\exists \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in [0,1]^n$ such that

$$f(\bar{\mathbf{x}}+\mathbf{dx}) - f(\bar{\mathbf{x}}) \quad = \quad \bigtriangledown f(\bar{\mathbf{x}}+\boldsymbol{\lambda}_1\cdot\mathbf{dx})\cdot\mathbf{dx} \tag{4a}$$

$$= \quad \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} \quad + \quad 0.5\mathbf{dx}'\mathrm{Hf}(\mathbf{x}+\boldsymbol{\lambda}_2\cdot\mathbf{dx})\cdot\mathbf{dx} \tag{4b}$$

The reason I call this the "global Taylor" theorem is that there's no restriction on the magnitude of $\mathbf{dx}$. By contrast, when we get to "local Taylor" we'll see that the theorem holds only for $\mathbf{dx}$'s that are sufficiently small. There's a close analog here between the difference between a global max and a local max.

The last term in the theorem is called *the remainder term*; it differs from the other terms in the expansion because it is evaluated at some point on the line-segment between $\bar{x}$ and $\bar{x}+dx$. A priori, we have no idea of the value of λ. So how can this theorem be of any use to us?

There are two cases in particular in which it is extremely useful. The second is by far the more important.

(1) if f is an $k+1$'th order polynomial, then the remainder term independent of its first argument, so that for the k'th order Taylor expansion, the equality holds for *all* values of λ. E.g., let $k=1$ and consider the quadratic $a+bx+cx^2$. In this case, in expression (3), the term $f^{(k+1)}(\bar{x}+\lambda dx)=2c$, for all x, λ, dx and the remainder term, $\frac{f^{(k+1)}(\bar{x}+\lambda dx)dx^{k+1}}{(k+1)!}$, is simply cdx^2, for all λ. In summary, when you are working with polynomial functions, if you take enough derivatives at a single point x, you can recover the *exact* value of the function at an *arbitrary* point in the domain, $x+dx$.

(2) if the Hessian of f is *globally* a definite (or semi-definite) matrix, then you always know the *sign* of the remainder term, even though you don't know the value of λ. E.g., if f is a strictly concave, twice differentiable function, then $\mathbf{Hf}(\cdot)$ is everywhere negative semi-definite, i.e., the remainder term $0.5\mathbf{dx}'\mathbf{Hf}(\cdot)\mathbf{dx}$ is *always* nonpositive. Hence we know that $f(\bar{\mathbf{x}}+\mathbf{dx})-f(\bar{\mathbf{x}})$ is *always* weakly less than the differential approximation $f(\bar{\mathbf{x}})+\bigtriangledown f(\bar{\mathbf{x}})\cdot\mathbf{dx}$. This fact has significant implications: it means that when $\mathbf{Hf}(\cdot)$ is everywhere negative semi-definite, the first order condition $\bigtriangledown f(\bar{\mathbf{x}})=0$ is both necessary *and* sufficient for $\mathbf{x}$ to be a (weak) *global* maximum.

Returning to functions where λ really does make a difference, you might think that if the $(k+1)$'th derivative of f at $\bar{\mathbf{x}}$ were really huge, then the remainder term, which is determined by this term, would be really huge also, and thus mess up your approximation in the sense that the remainder term would be much larger in absolute value than the terms that have been written out explicitly. However, if an important caveat is satisfied, it turns out that *any* order of the Taylor expansion will be "good enough"—in the sense of determining the *sign* of the left hand side—provided that the length of $\mathbf{dx}$ is small enough. The caveat is that the $k'th$ terms in the approximation must be non-zero. When $n>1$ and k is odd—in particular when $k=1$—whether or not this caveat is satisfied depends on the direction of $\mathbf{dx}$. Indeed, if the domain of f is $\mathbb{R}^n$, $n>1$, it will *always* fail to be satisfied for *some* direction(s) $\mathbf{dx}$ (since there always exists $\mathbf{dx}$ such that $\bigtriangledown f(\mathbf{x})\cdot\mathbf{dx}=0$).

Theorem Taylor-Young's Theorem (Local Taylor): Let $f:\mathbb{R}\to\mathbb{R}$ be $(K+1)$ times continuously differentiable and fix $\bar{x}\in\mathbb{R}$. For any $0\leq k\leq K$, if $\underbrace{f^{(k)}(\bar{x})\neq 0}_{k\text{'th term non-zero caveat}}$ there exists $\epsilon>0$ s.t. $\forall dx\in\mathbb{R}$ with $0<|dx|<\epsilon$,

$$\Big|\underbrace{f'(\bar{x})dx+\frac{f''(\bar{x})dx^2}{2}+...+\frac{f^{(k)}(\bar{x})dx^k}{k!}}_{k\text{'th order Taylor expansion}}\Big| \quad > \quad \Big|\underbrace{\frac{f^{(k+1)}(\bar{x}+\lambda dx)dx^{k+1}}{(k+1)!}}_{\text{Remainder Term}}\Big| \tag{5a}$$

Let $f:\mathbb{R}^n\to\mathbb{R}$ be thrice continuously differentiable and fix $\bar{\mathbf{x}}\in\mathbb{R}^n$. For any any $\mathbf{dx}\in\mathbb{R}^n$ $\underbrace{\text{such that } \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}\neq 0}_{1st\text{ term non-zero caveat}}$ $\exists\epsilon>0$ s.t. if $||\mathbf{dx}||<\epsilon$,

$$|\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}| \quad > \quad |0.5\mathbf{dx}'\mathbf{Hf}(\bar{\mathbf{x}}+\boldsymbol{\lambda}_2\cdot\mathbf{dx})\mathbf{dx}| \tag{5b}$$

(5c)

Similarly, if $\underbrace{\mathbf{dx}'\mathbf{Hf}(\bar{\mathbf{x}})\mathbf{dx}\neq 0}_{2nd\text{ term non-zero caveat}}$, $\exists\epsilon>0$ s.t. if $0<||\mathbf{dx}||<\epsilon$

$$|\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}+0.5\mathbf{dx}'\mathbf{Hf}(\bar{\mathbf{x}})\mathbf{dx}| \quad > \quad |\text{remainder term}| \tag{5d}$$

It should be emphasized that if $k > 1$, then the theorem has content even if the first $k-1$ terms *are* zero, provided the k'th term isn't. To see the significance of the "k'th term nonzero caveat" consider an unconstrained optimum of the function. In this case, the first order term in the Taylor series is necessarily zero, since $f'(x)$ or $\bigtriangledown f(\mathbf{x})$ is necessarily zero, demonstrating that for this important case, the "$k = 1$" version of Local Taylor is necessarily useless to us.

The intuition for Local Taylor is clearest when f maps $\mathbb{R}$ to $\mathbb{R}$ and (a) the first thru $k-1$'th order terms are zero; and (b) the k'th order term, $\frac{f^{(k)}dx^k}{k!}$, is nonzero. By the Taylor Lagrange theorem, we have in this case that for some $\lambda \in [0,1]$,

$$\begin{aligned} f(x+dx) - f(x) &= \underbrace{\frac{f^{(k)}(\bar{x})dx^k}{k!}}_{k\text{'th order Taylor expansion}} + \underbrace{\frac{f^{(k+1)}(x+\lambda dx)dx^{k+1}}{(k+1)!}}_{\text{Remainder term}} \\ &= \frac{dx^k}{k!}\left(f^{(k)}(\bar{x}) + dx\frac{f^{(k+1)}(x+\lambda dx)}{(k+1)}\right) \end{aligned} \tag{6}$$

Consider $dx > 0$. If dx is sufficiently small, then the first term in parentheses is going to dominate the second term, and $(f(x+dx) - f(x))$ is going to have the same sign as $f^{(k)}(x)$.

When k is odd, there is a striking difference in the applicability of Local Taylor depending on whether the domain of f is $\mathbb{R}$ or $\mathbb{R}^n$. For example, set $k = 1$. In the former case, the condition is that $f'(x) \neq 0$; in the latter it is that $\bigtriangledown f(\mathbf{x})\mathbf{dx} \neq 0$. I.e., in the latter case, it's possible that the analog of $f'(x) \neq 0$ is satisfied—i.e., $\bigtriangledown f(x) \neq 0$—but the "1st term non-zero caveat" for the theorem fails. Thus, when $n = 1$, whether or not the caveat $f'(\mathbf{x}) \neq 0$ is satisfied depends only on $\mathbf{x}$. But when $n > 1$, whether or not the caveat $\bigtriangledown f(\mathbf{x})\mathbf{dx} \neq 0$ is satisfied depends on both $\mathbf{x}$ *and* the direction of $\mathbf{dx}$. Indeed, if the domain of f is $\mathbb{R}^n$, $n > 1$, the caveat will *always* fail to be satisfied for *some* direction(s) $\mathbf{dx}$ (since there always exists some $\mathbf{dx}$ such that $\bigtriangledown f(\bar{\mathbf{x}}) \cdot \mathbf{dx} = 0$).

Relationship between the Two Taylor Theorems: We'll consider the scalar version:

- **Global:**

$$f(\bar{x}+dx) - f(\bar{x}) = f'(\bar{x})dx + ... + \frac{f^{(k)}(\bar{x})dx^k}{k!} + \frac{f^{(k+1)}(\bar{x}+\lambda dx)dx^{k+1}}{(k+1)!}$$

- **Local:** If dx is "small enough" and the "k'th term nonzero caveat" is satisfied:

$$\left|f'(\bar{x})dx + ... + \frac{f^{(k)}(\bar{x})dx^k}{k!}\right| > \left|\frac{f^{(k+1)}(\bar{x}+\lambda dx)dx^{k+1}}{(k+1)!}\right|$$

- **Implication:**

$$\begin{aligned} f(\bar{x}+dx) - f(\bar{x}) &= \underbrace{f'(\bar{x})dx + ... + \frac{f^{(k)}(\bar{x})dx^k}{k!}}_{k\text{'th expansion}} + \underbrace{\frac{f^{(k+1)}(\bar{x}+\lambda dx)dx^{k+1}}{(k+1)!}}_{\text{Remainder}} \\ \left|k\text{'th expansion}\right| &> \left|\text{Remainder}\right| \quad \text{implies} \\ \text{sgn}(f(\bar{x}+dx) - f(\bar{x})) &= \text{sgn}\left(f'(\bar{x})dx + ... + \frac{f^{(k)}(\bar{x})dx^k}{k!}\right) \end{aligned}$$

- **Conclusion:** If $|dx|$ is small, can sign the LHS without knowing *anything* about Remainder

Illustration of Taylor's theorem for $k = 1$: The purpose of this example is to illustrate, that

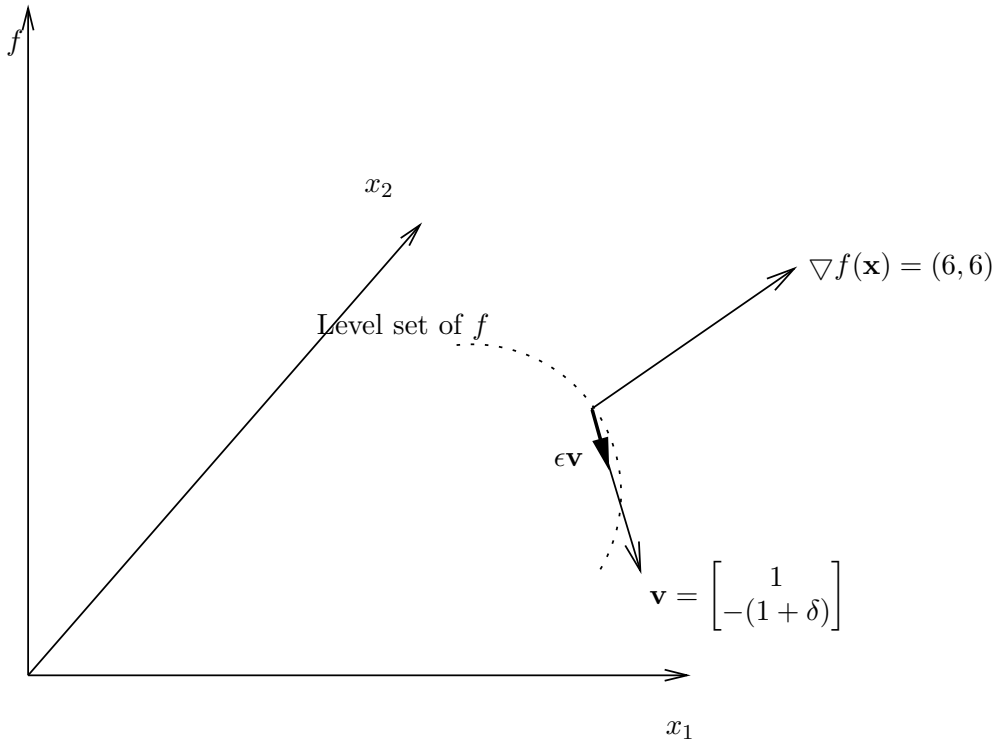

FIGURE 7. 1st order approx "works" if $\epsilon \approx 0$

(1) provided the direction of movement $\mathbf{dx}$ isn't orthogonal to the gradient, in which case the caveat of Taylor's theorem would fail for $k = 1$, then the sign of the linear approximation to the change in f will agree with the sign of the true change in f, *provided that* the magnitude of the shift $\mathbf{dx}$ is sufficiently small.
(2) whenever there exists $\mathbf{dx}$ such that $\mathbf{dx}'Hf(\mathbf{x})\mathbf{dx} \neq 0$, there will *never* exist an $\epsilon > 0$ such that for *any* $\mathbf{dx}$ with $||\mathbf{dx}|| < \epsilon$, the sign of the linear approximation to the change in f will agree with the sign of the true change in f.

this implies that in order to be sure that the sign of the first order Taylor approximation agrees with the actual sign of $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}})$, you must *first* choose the direction $\mathbf{h}$ and *then* determine the supremum of the lengths of the vector $\mathbf{dx} = \epsilon\mathbf{h}$ for which the first order Taylor approximation has this property. Suppose that $f(x) = 3x_1^2 + 3x_2^2$, so that $\bigtriangledown f(x) = \begin{bmatrix} 6x_1 \\ 6x_2 \end{bmatrix}$ and $\mathrm{Hf}(x) = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}$. When $\bar{\mathbf{x}} = (1, 1)$, then

$$\begin{aligned} & f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \\ = \quad & \bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx} + \tfrac{1}{2}\mathbf{dx}'\mathrm{Hf}(\bar{\mathbf{x}})\mathbf{dx} \\ = \quad & \begin{bmatrix} 6 & 6 \end{bmatrix}\mathbf{dx} + \tfrac{1}{2}\mathbf{dx}'\begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}\mathbf{dx} \\ = \quad & 6\left(dx_1 + dx_2 \quad + \quad \tfrac{1}{2}(dx_1^2 + dx_2^2)\right) \end{aligned}$$

Notice that the entire Taylor expansion has exactly two terms, so that instead of an approximation sign in the display above, you have an equality. That is, when $k = 2$, there *is* no remainder

term. Next note that if $\mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, and $\mathbf{dx} = \epsilon\mathbf{v}$, for some $\epsilon \in \mathbb{R}$, then the *first* term in the Taylor expansion is zero, while the second is $6\epsilon^2$. Thus the first term in the Taylor expansion is dominated in absolute value by the second, regardless of the length of ϵ. Fortunately, however, this isn't a counter-example to Local Taylor since in the direction $\mathbf{v}$, the first order term in the Taylor expansion is zero, so that when $k = 1$, the "first term nonzero caveat" is not satisfied.

Now fix an arbitrary $\delta > 0$ and consider $\mathbf{v} = \begin{bmatrix} 1 \\ -(1+\delta) \end{bmatrix}$. With this modification, the first term of the Taylor expansion in the direction $\mathbf{v}$ is $-6\delta < 0$. Thus, the caveat in Taylor's theorem *is* satisfied for $k = 1$, and so the theorem had better work for this k. Indeed, we'll show that there exists $\bar{\epsilon} > 0$ such that if $\epsilon < \bar{\epsilon}$ and $\mathbf{dx} = \epsilon\mathbf{v}$, then $|\bigtriangledown f(\bar{\mathbf{x}})\mathbf{dx}| > |\frac{1}{2}\mathbf{dx}'\mathbf{Hf}(\bar{\mathbf{x}})\mathbf{dx}|$, or, in other words, the sign of $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}})$ will agree with the sign of $-6\epsilon\delta$.

Let $\mathbf{dx} = \epsilon\mathbf{v}$, for $\epsilon > 0$. Observe that the first term in the Taylor expansion is negative ($-6\delta\epsilon < 0$), while

$$\begin{aligned} f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) &= 6\left(dx_1 + dx_2 \quad + \quad \frac{1}{2}(dx_1^2 + dx_2^2)\right) \\ &= 6\left(-\epsilon\delta \quad + \quad \frac{1}{2}[\epsilon^2 + \epsilon^2(1+\delta)^2]\right) \\ &= 6\epsilon\Big(-\delta \quad + \quad \epsilon[\delta + 1 + \delta^2/2]\Big) \\ &= 6\epsilon\delta\Big(-1 \quad + \quad \epsilon[\underbrace{1 + 1/\delta + \delta/2}_{\to\infty \text{ as } \delta\to 0}]\Big) \end{aligned}$$

Note that if $\epsilon > 0$ is, say greater than unity, then $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}})$ is *positive*. On the other hand, provided that $\epsilon < \bar{\epsilon} = \frac{1}{1+1/\delta+\delta/2}$ then $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}})$ will be negative, just like the first term in the Taylor expansion!

I'll now give an alternative, purely graphical, explanation of why it is impossible to pick an $\epsilon > 0$ such that the first order Taylor expansion will give the correct sign for *all* vectors of length not greater than ϵ. In Fig. 8, the circle centered at $\mathbf{x}$ is of radius ϵ. Note that if you consider a vector of length ϵ that points into one of the two dashed cones emanating from $\mathbf{dx}$ (for example $\mathbf{v}^{\text{bad}}$), then it will pass thru the lower contour set of f corresponding to $\mathbf{x}$ and out the other side into the *upper* contour set. On the other hand, since $\mathbf{v}^{\text{bad}}$ makes an obtuse angle with $\bigtriangledown f(\mathbf{x})$, the differential $\bigtriangledown f\mathbf{xdx}$ is negative, i.e., gives the incorrect sign. For a vector such as $\mathbf{v}^{\text{good}}$, which lies outside the dashed cones, the sign of the differential is the same as the sign of the actual difference $f(\mathbf{x} + \mathbf{dx}) - f(\mathbf{x})$.

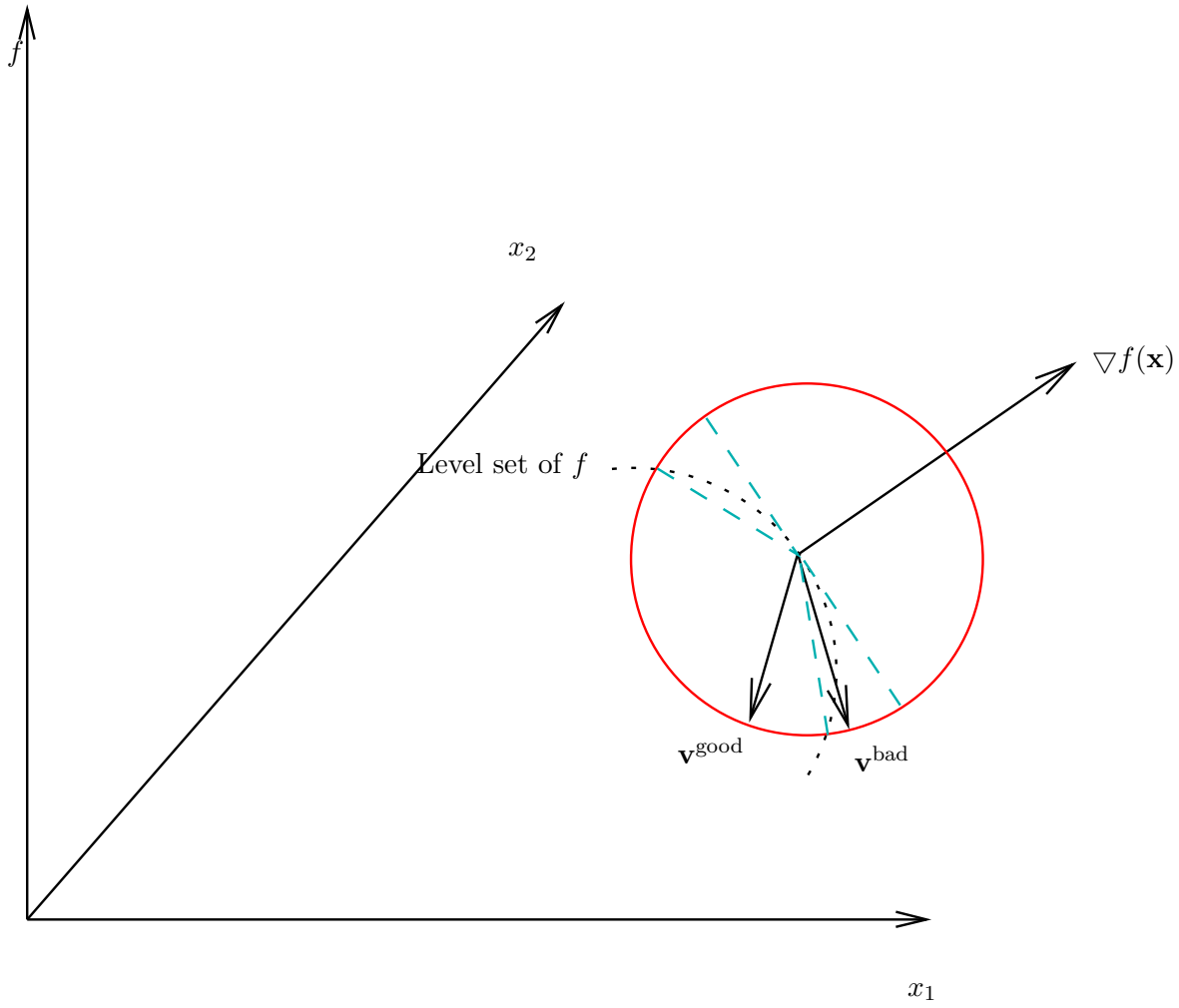

FIGURE 8. No $\epsilon > 0$ works for all directions