The marginal rate of substitution

Following the explanation in the text, you might expect that if two goods each exhibit diminishing marginal utility, then the marginal rate of substitution between them will also be diminishing. As we will show below, the marginal rate of substitution is equal to the ratio of the marginal utilities of the two goods, so this hypothesis seems reasonable. Unfortunately, this hypothesis is not generally true if the marginal utility of one good depends upon the amount consumed of the other good. For example, an increase in your butter consumption may decrease your marginal utility of margarine, or an increase in your coffee consumption may increase your marginal utility of cream. This note investigates the conditions under which the marginal rate of substitution is decreasing, as hypothesized in the text.

Consider a general utility function for two goods X and Y: $U = U(X,Y)$ with $\frac{\partial U(X,Y)}{\partial X} = U_X > 0$ and $\frac{\partial U(X,Y)}{\partial Y} = U_Y > 0$. That is, we assume that the marginal utility of both goods is positive, but make no assumptions regarding the slope of marginal utility. (For convenience, we have adopted the conventional notation of U_X and U_Y as the partial derivatives of U with respect to X and Y.) By definition, the marginal rate of substitution (MRS) is equal to the slope of an indifference curve: for a given change in X, the amount by which Y must change so as to keep utility constant. Mathematically, MRS = $\left.\frac{\mathrm{d}Y}{\mathrm{d}X}\right|_{\text{U = constant}}$. To find this derivative, we begin by taking the total differential of the utility function $\mathrm{d}U = U_X\mathrm{d}X + U_Y\mathrm{d}Y$ and then set $\mathrm{d}U$ = zero to hold utility constant: $0 = U_X\mathrm{d}X + U_Y\mathrm{d}Y$. A bit of rearranging gives the desired result: $\left.\frac{\mathrm{d}Y}{\mathrm{d}X}\right|_{\text{U = constant}} = -\frac{U_X}{U_Y}$. In words, the marginal rate of substitution is equal to the ratio of the marginal utilities. For convenience, let us define MRS = $-\frac{U_X}{U_Y} = R(X, Y)$. That is, U_X and U_Y are each functions of X and Y, so their ratio is also a function of X and Y. Call this ratio $R(X, Y)$.

Convexity of an indifference curve requires that its slope be diminishing in absolute value; since the slope is negative, this means that the slope must be increasing in X. In symbols, we require that $\frac{\mathrm{d}R(X,Y)}{\mathrm{d}X} > 0$. As before, we begin our task by taking the total differential of $R(X, Y)$: $\mathrm{d}R = R_X\mathrm{d}X + R_Y\mathrm{d}Y$ where R_X and R_Y are the partial derivatives of R with respect to X and Y. Dividing through by $\mathrm{d}X$, we obtain the result, $\frac{\mathrm{d}R}{\mathrm{d}X} = R_X + R_Y\frac{\mathrm{d}Y}{\mathrm{d}X}$. Here we make use of the fact that $\frac{\mathrm{d}Y}{\mathrm{d}X} = R(X, Y) = -\frac{U_X}{U_Y}$ and substitute to find $\frac{\mathrm{d}R}{\mathrm{d}X} = R_X - R_Y\frac{U_X}{U_Y}$.

We are not finished, however, for we must find R_X and R_Y. First, $R_X = \frac{\partial\left(-\frac{U_X}{U_Y}\right)}{\partial X}$ so we will need to use the division rule: $\frac{\partial\left(-\frac{U_X}{U_Y}\right)}{\partial X} = -\frac{U_Y\left(\frac{\partial U_X}{\partial X}\right) - U_X\left(\frac{\partial U_Y}{\partial X}\right)}{(U_Y)^2}$. Likewise, $R_Y = \frac{\partial\left(-\frac{U_X}{U_Y}\right)}{\partial Y} =$

$-\dfrac{U_Y\left(\dfrac{\partial U_X}{\partial Y}\right)-U_X\left(\dfrac{\partial U_Y}{\partial Y}\right)}{(U_Y)^2}$. The terms in parentheses are actually second derivatives of the utility function: $\dfrac{\partial U_X}{\partial X}=\dfrac{\partial^2 U}{\partial X^2}=U_{XX}$ (to continue our previous notational shorthand.) Likewise, $\dfrac{\partial U_Y}{\partial Y}=\dfrac{\partial^2 U}{\partial Y^2}=U_{YY}$ and $\dfrac{\partial U_X}{\partial Y}=\dfrac{\partial U_Y}{\partial X}=\dfrac{\partial^2 U}{\partial X \partial Y}=U_{XY}$. Making these substitutions into the expression for $\dfrac{dR}{dX}$, we find that $\dfrac{dR}{dX}=-\dfrac{U_Y U_{XX}-U_X U_{XY}}{(U_Y)^2}+\left(\dfrac{U_X}{U_Y}\right)\dfrac{U_Y U_{XY}-U_X U_{YY}}{(U_Y)^2}$. As messy as this looks, it can be simplified somewhat by multiplying and dividing the first term by U_Y to get a common denominator and then rearranging: $\dfrac{dR}{dX}=\dfrac{2U_X U_Y U_{XY}-(U_Y)^2 U_{XX}-(U_X)^2 U_{YY}}{(U_Y)^3}$. Finally!

Recall that the condition we seek is that $\dfrac{dR}{dX}>0$. The denominator is positive by our assumption of positive marginal utility for each good, so we only require that the numerator be positive. Now diminishing marginal utility of each good is assured if U_{XX} and U_{YY} are negative, in which case the last two terms of $\dfrac{dR}{dX}$ are positive. But what of the first term? It is certainly possible that U_{XY} is sufficiently negative (an increase in consumption of X reduces the marginal utility of good Y) so that the first term swamps the positive effect of the last two. In other words, it is possible that the indifference curve is not convex despite the fact that each good has diminishing marginal utility.

Likewise, $\dfrac{dR}{dX}$ may be positive even if the last two terms are negative. This possibility requires that the first term be sufficiently positive ($U_{XY}>0$, implying that increasing consumption of X increases the marginal utility of Y) so that it swamps the effect of the last two terms. That is, the indifference curve can be convex even if both goods exhibit <u>increasing</u> marginal utility! We are left with the uneasy conclusion that diminishing marginal utility is neither a sufficient nor a necessary condition for a diminishing marginal rate of substitution (convex indifference curves) and we must therefore simply state this as an assumption, albeit a reasonable one: The MRS between any pair of goods X and Y is diminishing.