Linear Programming has the virtue of solving very large problems relatively quickly. Not surprisingly there are applications of LP to Big Data. The first one I want to discuss is Compressed Sensing. We have several faculty with expertise (O. Yilmaz, Y. Plan). There is a wikipedia page that gives an overview.

The main idea is that in some problem domains, a data point ($\mathbf{b}$ below) can be expressed as a linear combination ($\mathbf{x}$ below) of a *few* 'basis' elements when the 'basis' itself is very large. For example in understanding a small piece of an image, one can imagine it can be represented using a small number of pictorial elements chosen from some large set of possibilities. There is plenty of experimental evidence that the 'basis' would consist of various possibilities for a black/white edge in the small image as well some few choices for colour and so only a small basis would be required to reperesent the picture well. In the problem below, A is an $m \times n$ matrix with $m << n$.

$$\begin{array}{lrcl} \max & ||\mathbf{x}||_1 & & \\ & A\mathbf{x} & = & \mathbf{b} \end{array}.$$

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$. Here we define the L_1 *norm* $||\mathbf{x}||_1$ as

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|.$$

There are other reasonable choices for the objective function. The typical L_2 norm (square root of sum of squares) is not so helpful here. If we believe that our solution $\mathbf{x}$ should be sparse with few non zero terms, then it may make sense to minimize the number of non zeros in $\mathbf{x}$ (sometimes called the L_0 norm of $\mathbf{x}$) but Candès and others have shown that it likely sufficient to use the L_1 norm instead. Minimizing the L_1 norm is computationally tractable because we can express it using Linear Programming as we shall see below.

Another application of big data is to use the massive data collected at websites to try and make profitable predictions. Imagine we have m customer and n characteristics (e.g. time spent at website, previous search history, time of day etc). It is a little harder to argue that $m << n$ but certainly we have lots of data to choose many sharacteristics. The right hand side values could be the money spent at the website. We are interested in determining which characteristics might predict higher spending. Choice of where to place ads may be relevant here for the company.

We solve the Mathematical problem by minimizing a function z which is constrained as being larger than the L_1 norm of $\mathbf{x}$

One way is:

$$\begin{array}{rcl} z_1 & \geq & x_1 \\ z_1 & \geq & -x_1 \\ z_2 & \geq & x_2 \\ z_2 & \geq & -x_2 \\ & \vdots & \\ z_n & \geq & x_n \\ z_n & \geq & -x_n \\ z & = & z_1 + z_2 + \cdots + z_n \end{array}$$

These $2n$ inequalities force $z \geq \sum_{i=1}^{n} |x_i|$ and so, if we use z as the objective function and we will be minimizing z, the solution will have $z = \sum_{i=1}^{n} |x_i|$. We are trying to make $\mathbf{x}$ small, in the L_1 norm sense, to correspond to an idea of compressed sensing.

You might note that since $m < n$, then the matrix has matrix rank quite a bit smaller than n and hence there are many solutions $\mathbf{z}$ to $A\mathbf{z} = \mathbf{0}$ (from MATH 221 or elsewhere the dimension of the subspace formed by such z is at least $n-m$). If we are lucky (?) enough to have one solution to $A\mathbf{x} = \mathbf{b}$, then for each such $\mathbf{z}$ we get a solution $A(\mathbf{x}+\mathbf{z}) = A\mathbf{x} + A\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{0}$ and hence many solutions to $A\mathbf{x} = \mathbf{b}$. Given somewhat random data it would be likely that a solution with equality can be found (it is equivalent to asking if $\mathbf{b}$ is in the columns space of A). We wish to choose an $\mathbf{x}$ which is the smallest one in some sense. There is evidence that minimizing the L_1 norm will also result in a sparsest solution, namely the fewest number of non zero entries in $\mathbf{x}$.

Applications of this continue to grow. For example sharpening an image that has been corrupted by noise (i.e. it is fuzzy). We can attempt to express the given image (well only a small portion at a time) as a linear combination of blocks which have been shown to yield sharp images.

It might be helpful to indicate why the following two problems are equivalent in that the objective function values at optimality are equal.

$$L_1 \text{ problem} \qquad \begin{array}{rrcl} \min & ||\mathbf{x}||_1 & & \\ & A\mathbf{x} & = & \mathbf{b} \end{array}.$$

$$LP: \qquad \begin{array}{rrcc} \min & z & & \\ & A\mathbf{x} & = & \mathbf{b} \\ & z & = & z_1 + z_2 + \cdots z_n \\ & z_1 & \geq & x_1 \\ & z_1 & \geq & -x_1 \\ & z_2 & \geq & x_2 \\ & z_2 & \geq & -x_2 \\ & \vdots & & \vdots \\ & z_n & \geq & x_n \\ & z_n & \geq & -x_n \end{array}.$$

Let a be the value of the objective function at optimality in the L_1 problem and let b be the value of the objective function at optimality in the LP problem

If $\mathbf{x}^*$ is an optimal solution to the L_1 problem then we can obtain a feasible solution to the LP by using $\mathbf{x} = \mathbf{x}^*$ and $z_i = |x_i|$ for $i = 1, 2, \ldots, n$ and $z = z_1 + z + 2 + \cdots + z_n$. Moreover the values of the objective function are equal. Thus $b \leq a$ since the LP has a solution of value a.

If $\mathbf{x}^*$ are the values for $\mathbf{x}$ in an optimal solution to the LP, then we have feasibility in the L_1 problem with $\mathbf{x} = \mathbf{x}^*$. Since $z_i \geq x_i$ and $z \geq -x_i$ (in the LP) then $z_i \geq |x_i$ and so the obejctive function value for the LP is at least the value for the L_1 problem. Thus $a \leq b$.

Putting these together, means that $a = b$ and so optimal solutions to the two problems have the same objective function value.

Donoho, David L. (2006). "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution", Communications on Pure and Applied Mathematics. 59 (6): 797–829.