Optimization $\equiv$ Nonlinear System

An optimization problem can be considered equivalent to solving $F(x) = 0$ with $F(x) = \nabla f(x)$. Such an $F(x)$ has special structure which we can exploit. For a given $f(x), \;\; x^{(r)} \in \Re^n$ and an arbitrary constant vector $u \in \Re^n$ define $g(t), \;\; g : \Re \to \Re$ by,

$$g(t) \equiv f(x^{(r)} \; + \; tu).$$

It then follows that after differentiating wrt t,

$$\frac{dg}{dt} \equiv g'(t) = (\frac{\partial f}{\partial x_1} u_1 + \frac{\partial f}{\partial x_2} u_2 \cdots + \frac{\partial f}{\partial x_n} u_n) = (\nabla f)^T u.$$

This expression describes how f changes in the 'direction' u. In particular, if $\nabla f(x)^T u < 0$ then f decreases in that direction and u is called a 'descent direction'. This leads to the method of steepest descent: where we choose $u = -\nabla f$ (to obtain $g'(t) = \nabla f^T u = -\|\nabla f\|_2^2$), and determine $\bar{t} \in \Re$ to minimise $g(t) = f(x^{(r)} - t\nabla f)$.

Steepest Descent

-guess $x^{(0)} = (x_1^{(0)}, x_2^{(0)} \cdots x_n^{(0)})^T$

-<u>for</u> $r = 0, 1 \cdots$ <u>until</u> satisfied <u>do</u>:

$u^r = \nabla f(x^{(r)}) \;\; (\equiv F(x^{(r)}))$

-<u>if</u> $u^r \approx \underline{0}$ <u>then</u> signal convergence

-<u>else</u>

-find $\bar{t}$ such that $g(t)$

$\equiv f(x^{(r)} - tu_r)$ is minimum

-$x^{(r+1)} = x^{(r)} - \bar{t}u^r$

-<u>end</u>

<u>end</u>

Care must be taken to ensure the stopping criteria of the inner iteration is consistent with that of the outer iteration.

Observations

- With this approach there is only a $1D$ line search on each iteration and any scalar nonlinear equation solver can be used (eg., Bisection, Newton or Secant). We will always observe a decrease $f(x^{(r+1)}) < f(x^{(r)}) \cdots < f(x^{(0)})$. One can prove that the sequence will always converge but convergence may be slow.
- Optimization methods can also be used to solve Nonlinear systems. That is, we can interpret a system of nonlinear equations as a special case of an optimization problem. To see this, consider the nonlinear system $F(x) = 0$ and define $h : \Re^n \to \Re$ by,

$$h(x) \equiv \|F(x)\|_2^2 = \sum_{i=1}^{n} f_i^2(x).$$

Clearly,

$$F(\alpha) = \underline{0} \Leftrightarrow h(\alpha) \text{ is minimum.}$$

Optimization to Solve $F(x) = 0$

With $F \equiv (f_1, f_2 \cdots f_n)^T$,

$$\nabla h(x) = (\frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_2} \cdots \frac{\partial h}{\partial x_n})^T.$$

where

$$\frac{\partial h}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{i=1}^{n} f_i^2(x) = 2 \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_j} f_i = 2 \left(W^T F \right)_j ,$$

where W is the Jacobian matrix defined by,

$$W \equiv \frac{\partial F}{\partial x} \text{ whose } (i\ j)^{th} \text{entry is } \frac{\partial f_i}{\partial x_j}.$$

Therefore we have that $\nabla h(x) = 2W^T F$ and this is the zero vector only when $F = \underline{0}$ (unless W is singular).

An Example Problem

You are given the following two equations in two unknowns:

$$3x_1^2 - x_2^2 = 0, \quad 3x_1x_2^2 - x_1^3 - 1 = 0.$$

- Consider applying Newtons method for systems to this problem with initial guess, $(x_1^1, x_2^1) = (1.0, 1.0)$. What is the linear equation that defines the value of the first iterate, (x_1^2, x_2^2)?
- Solve this linear system and determine (x_1^2, x_2^2).
- A better initial iterate can be determined based on the 'steepest descent' direction from the same initial guess. That is, by letting $h(x_1, x_2) = f_1^2 + f_2^2$ (where f_1 and f_2 correspond to the above two equations), determine the steepest descent direction for $h(x_1, x_2)$ at the initial guess and describe how you could compute an alternate first iterate using this information.

Numerical ODEs

- Definition: A first-order ordinary differential equation is specified by:

$$y' = f(x, y), \quad \text{over a finite interval} \quad x \in [a, b].$$

- Note that a solution of this ODE, $y(x)$, is a function of one variable. When the solution depends on more than one variable (ie a multivariate function) it is called a partial differential equation – PDE. The term first-order refers to the highest derivative that appears in the equation.
- For ODEs the variable x is called the independent variable while y (which depends on x) is called the dependent variable. 'Solving' the ODE is interpreted as determining a technique for expressing y as a function of x in some explicit way.

ODEs-Mathematical Preliminaries

- A function $\Phi(x)$ is a solution of this ODE if $\Phi(x) \in C^1[a,b]$ and $\forall x \in [a,b]$ we have $\Phi^{'}(x) = f(x, \Phi(x))$. (Note that this condition is often easy to check or verify).
- For example the ODE,

$$y' = \lambda y,$$

has solutions $\Phi(x) = c\, e^{\lambda x}$ for any constant c since,

$$[c\, e^{\lambda x}]^{'} = \lambda c e^{\lambda x} = \lambda \Phi(x).$$

In particular this ODE does not have a unique solution but rather a whole family of solutions (characterized by the parameter c).

ODEs-Mathematical Preliminaries

- To determine a unique mathematical solution we must add an additional constraint. The most common way to do this is to prescribe the value of the solution at the initial point of the interval. That is we specify,

$$y(a) = y_0.$$

  –Definition: An ODE together with the initial conditions specifies an initial value problem for an ordinary differential equation (IVP for an ODE).

- Before we can attempt to approximate a solution to an IVP we must consider some essential mathematical questions:
  - Does a solution exist?
  - If a solution exists, is it unique?
  - Can the problem be solved analytically (ie. in closed form)?

IVPs - Existence/Uniqueness

- Definition: The function $f(x, y)$ satisfies a Lipschitz condition in y (ie, wrt its second argument) if $\exists L > 0$ such that $\forall x \in [a, b]$ and $\forall\, u, v$ we have

  $$|f(x, u) - f(x, v)| \le L|u - v|.$$

  In particular, if $f(x, y)$ has a continuous partial derivative with respect to y and this derivative is bounded for all y, then f satisfies a Lipschitz condition in y since,

  $$|f(x, u) - f(x, v)| = |\frac{\partial f}{\partial y}(x, \eta)|\; |u - v|,$$

  for some η between u and v.

IVPs - Existence/Uniqueness

- A typical Theorem:
  Let $f(x, y)$ be continuous for $x \in [a, b]$ and $\forall y$ and satisfy a Lipschitz condition in y, then for any initial condition y_0 the IVP,

$$y' = f(x, y), \quad y(a) = y_0, \quad \text{over } [a, b],$$

  has a unique solution, $y(x)$ defined for all $x \in [a, b]$.

Systems of ODEs

- Often one must deal with a system of n 'unknown' dependent variables of the form:

$$\begin{array}{rcl} y_1' & = & f_1(x, y_1, y_2, \cdots y_n), \\ y_2' & = & f_2(x, y_1, y_2, \cdots y_n), \\ \vdots & \vdots & \vdots \\ y_n' & = & f_n(x, y_1, y_2, \cdots y_n), \end{array}$$

with initial conditions all specified at the same point,

$$\begin{array}{rcl} y_1(a) & = & c_1, \\ y_2(a) & = & c_2, \\ \vdots & \vdots & \vdots \\ y_n(a) & = & c_n, \end{array}$$

Systems of ODEs (cont)

In vector notation, this system of IVPs can be written

$$Y' = F(x, Y), \quad Y(a) = Y_0,$$

where $Y(x) = [y_1(x), y_2(x), \cdots y_n(x)]^T, \; Y_0 = [c_1, c_2, \cdots c_n]^T$ and $F(x, Y)$ is a vector-valued function,

$$F(x, Y) = \begin{bmatrix} f_1(x, Y) \\ f_2(x, Y) \\ \vdots \\ f_n(x, Y) \end{bmatrix}.$$

The theory and the investigation of numerical methods that we present will be the same for systems as for scalar IVPs. In particular, the Theorem quoted above holds for systems.

Some Examples

- From Biology:
  A predator-prey relationship can be modeled by the IVP:

$$y_1' = y_1 - 0.1y_1y_2 + 0.02x$$

$$y_2' = -y_2 + 0.02y_1y_2 + 0.008x$$

  with

$$y_1(0) = 30, \quad y_2(0) = 20.$$

  Here $y_1(x)$ represents the 'prey' population at time x and $y_2(x)$ represents the 'predator' population at time x. The solution can then be visualized as a standard x/y solution plot or by a 'phase plane' plot. Figure 1 illustrates the solution to this system. We know that for different initial conditions solutions to this problem exhibit oscillatory behaviour as x increases.

Solution to PP problem

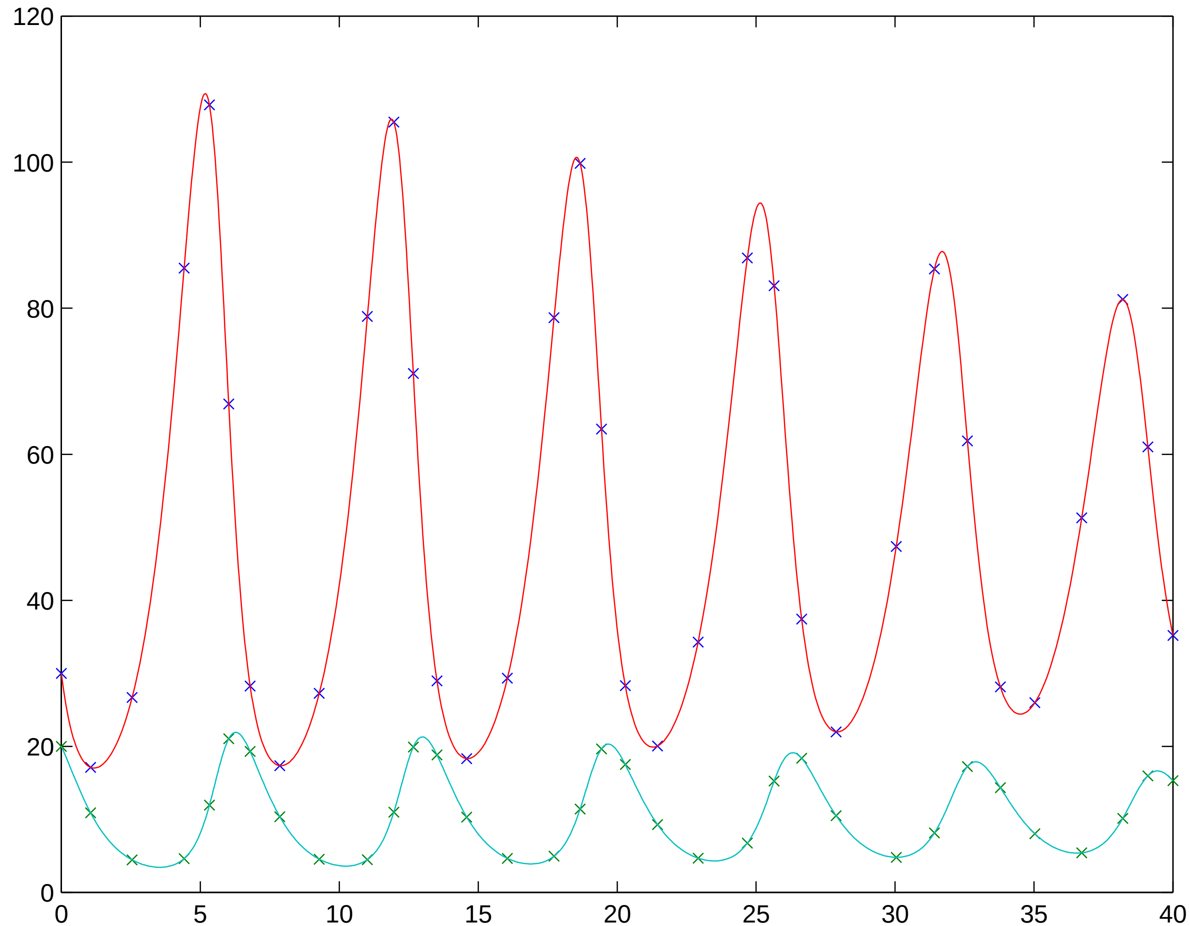

Figure 1. Solution plot for the Predator-Prey Problem.

Solution to PP problem

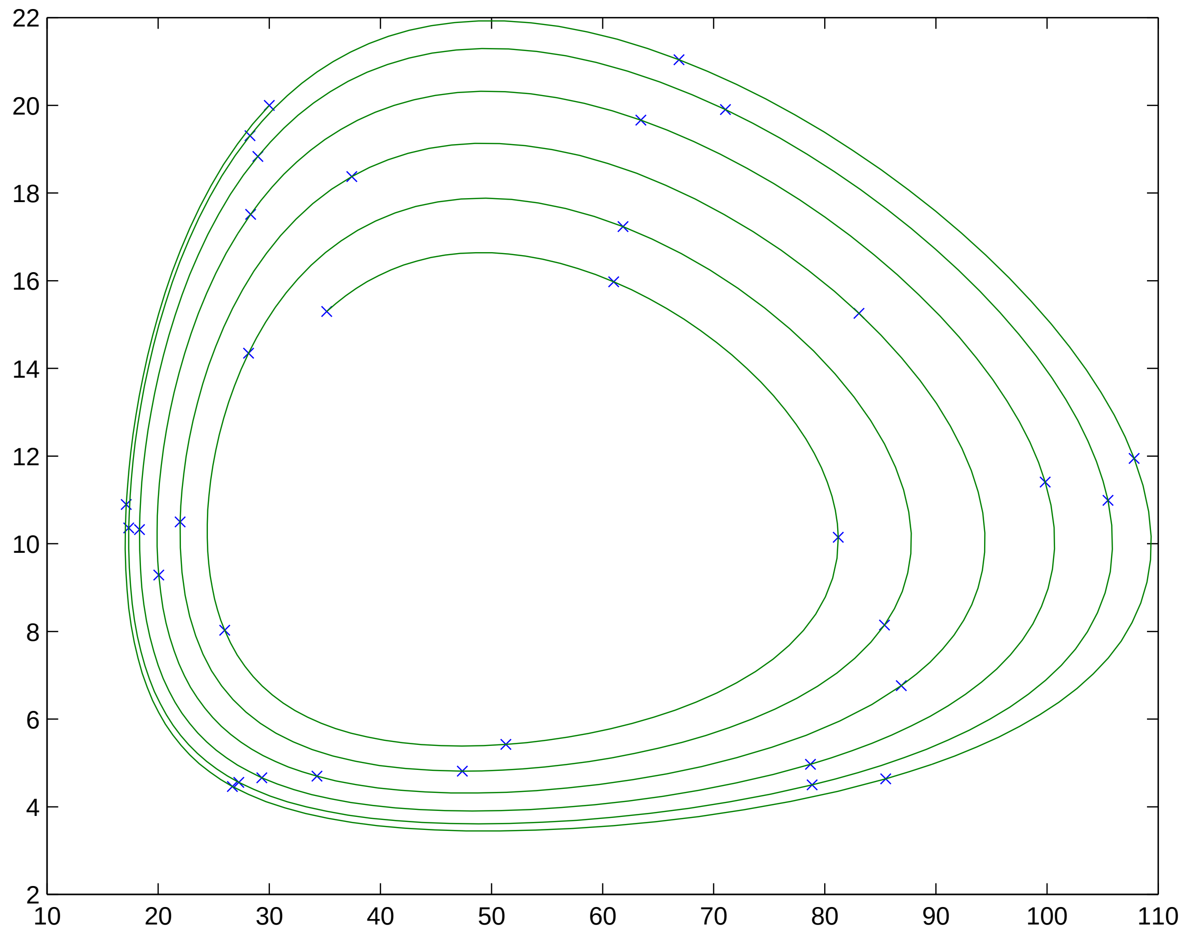

Figure 2. Phase Plane Plot for Predator-Prey Problem.

Application of IVPs

A biologist may be interested in whether the solutions to this equation are ‘almost periodic’ (in the sense that the difference between successive maximum is constant) and whether the local maxima approach a steady state exponentially.

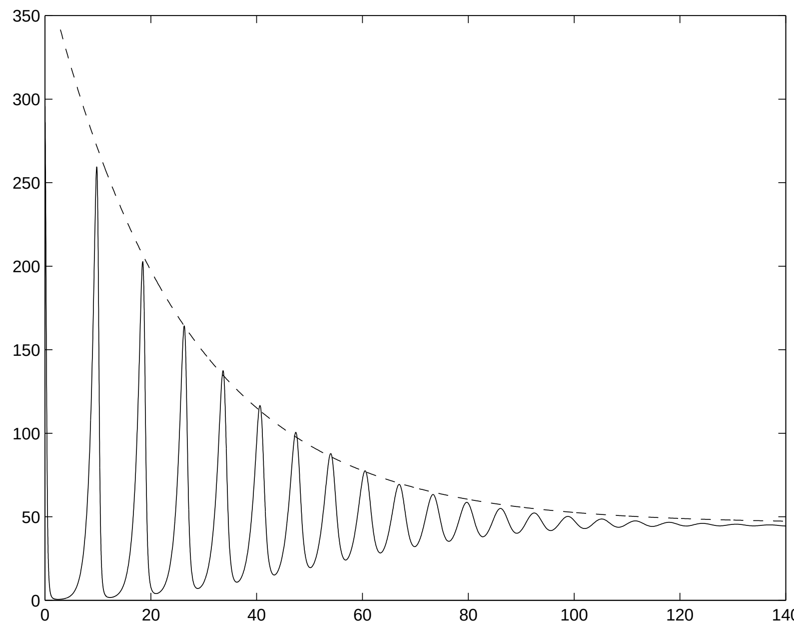

Fig. 3. Typical behaviour of prey population and decay to steady state.