TMA 4180 Optimeringsteori
Exam May 30, 2005
Solution with additional comments

PRELIMINARY VERSION

Note that this solution is more detailed than required for full score at the exam. The problems are only included for an easier reading.

Problem 1:

We consider the following problem:

$$\min_{x \in D \subset \mathbb{R}^n} \left\{ \frac{1}{2} x' C x \right\}, \tag{1}$$

where C is positive definite ($C > 0$). The feasible set D is defined by

$$D = \left\{ x \ ; \ a_i' x \geq \beta_i \text{ for some } i \in \{1, 2, \cdots, m\} \right\}, \tag{2}$$

where $a_j \in \mathbb{R}^n$, $j = 1, \cdots, m$. and $(\beta_1, \beta_2, \cdots, \beta_m)' \in \mathbb{R}^m$

(D is the union of the sets $\left\{ x \ ; a_j' x \geq \beta_i \right\}$, $i = 1, \cdots, m$). All redundant inequalities have been removed, and 0 is not contained in D.

(a) Explain how the problem may be solved. Is the solution unique, and if not, are all local minima also global minima? (Hint: A sketch may be helpful).

(b) Apply your method from (a) to solve the problem

$$\begin{gathered} \min \left\{ x^2 + 2y^2 + z^2 \right\} \\ x + 2y + 3z \geq 3, \text{ or} \\ x - y + z \geq 2, \text{or} \\ x + 2y - 3z \geq 1. \end{gathered} \tag{3}$$

Solution:

(a) The matrix is positive definite, and hence the level surfaces (contours in $\mathbb{R}^2$) are (hyper-) ellipses. However, the feasible domain will not be convex unless $m = 1$ (since redundant inequalities are removed). Recall that $\left\{ x \ ; a_j' x \geq \beta_i \right\}$ define closed half-spaces. It is illustrative to make a graph in $\mathbb{R}^2$, see Fig. 1. The problem is then to find the the points in Ω that are closest to the origin, when the distance is measured in the C-metric.

The problem is most easily solved by first considering the "elementary" sub-problem

$$\begin{gathered} \min_{x \in D \subset \mathbb{R}^n} \left\{ x' C x \right\}, \\ D_0 = \left\{ x \ ; \ a' x \geq \beta \right\} \end{gathered} \tag{4}$$

Of course, if $0 \in D_0$, the solution is simply $x = 0$, but this was excluded here. It is clear that since $C > 0$, the objective function $f(x) = x' C' x$ is strictly convex. Moreover, since also D_0 is convex, we always have (since $f(x) \to \infty$ when $|x| \to \infty$) a *unique* solution fulfilling

$$\nabla f(x^*) - \lambda a' = 0. \tag{5}$$

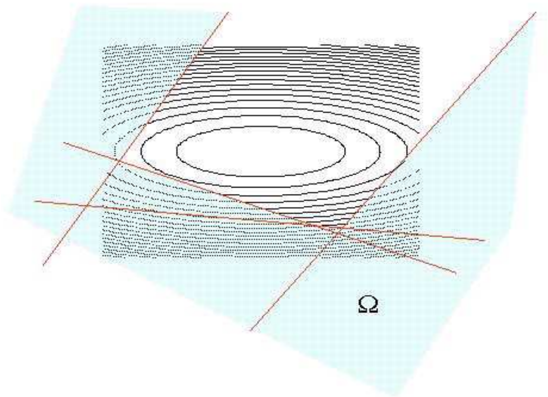

Figure 1: Contours of $x'Cx$ and constraints defining the feasible domain Ω. Note that even the complement of Ω may be unbounded.

Since $\nabla f(x^*)' = Cx^*$, we first solve $Cx_0^* = a$ and then adjust the length, $x^* = sx_0^*$, so that $a'x^* = \beta$. The solution is therefore

$$x^* = \frac{\beta}{a'C^{-1}a} C^{-1} a. \tag{6}$$

We have to solve this elementary problem for all inequalities, and the solutions x^* for which $x^{*\prime}Cx^*$ has the smallest value, are solutions to the full problem. Looking at the problem, it is easy to see that

- we always have solutions.
- two elementary solutions are never equal.
- there may be be up to m different optimal solutions.
- Elementary solutions are local minima if and only if they are members of only one half-space

(b) In this particular case, $C = \text{diag}\{1\ 2\ 1\}$ and $C^{-1} = \text{diag}\{1\ \frac{1}{2}\ \ 1\}$. The elementary solutions for the three inequalities are

$$\begin{aligned}
x^*_{(1)} &= \frac{3}{(1\ 2\ 3)\, C^{-1}\, (1\ 2\ 3)'} C^{-1} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \\
x^*_{(2)} &= \frac{2}{(1\ -2\ 3)\, C^{-1}\, (1\ -2\ 3)'} C^{-1} \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}, \\
x^*_{(3)} &= \frac{1}{(1\ 2\ -3)\, C^{-1}\, (1\ 2\ -3)'} C^{-1} \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix},
\end{aligned} \tag{7}$$

The corresponding objective values, $x^{*\prime}Cx^*$,

$$\begin{aligned} f_{(1)} &= \frac{3}{4}, \\ f_{(2)} &= \frac{1}{3}, \\ f_{(3)} &= \frac{1}{12}, \end{aligned} \tag{8}$$

and the unique solution (all objective values are different) is thus $x^* = x^*_{(3)} = \frac{1}{12}(1, 1, -3)'$.

Problem 2:

(a) What are the main ideas behind the Gauss-Newton method for non-linear least squares problems?

(b) Formulate the minimization of the Rosenbrock Banana Function

$$f(x) = 100\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 \tag{9}$$

as a least-square problem.

Carry out one step with the Gauss-Newton method, starting at $(0, 0)$. *It is enough to write down the equations, and not necessary to compute final numerical values.*

Solution:

(a) Gauss-Newton is a method for non-linear least square problems where we have a vector-valued function, $h(x) \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and we want to find the minimum of

$$f(x) = \frac{1}{2}\|h(x)\|_2^2 = \frac{1}{2}\sum_{i=1}^{m} h_i(x)^2 = \frac{1}{2}h(x)'h(x). \tag{10}$$

The Jacobian matrix of h is defined by

$$J(x) = \left\{\frac{\partial h_i}{\partial x_j}(x)\right\}_{\substack{i=1,\ldots,m,\\ j=1,\ldots,n}} = \begin{bmatrix} \nabla h_1(x) \\ \nabla h_2(x) \\ \vdots \\ \nabla h_m(x) \end{bmatrix}. \tag{11}$$

and

$$\nabla f(x)' = J'(x)h(x), \tag{12}$$

$$F(x) = \nabla^2 f(x) = J'(x)J(x) + \sum_{i=1}^{m} h_i(x)F_i(x). \tag{13}$$

The full Newton Method for this problem would have been

$$F(x_k)(x_{k+1} - x_k) = -\nabla f(x_k)', \;\; k = 0, 1, \cdots, \tag{14}$$

and the Gauss-Newton method uses the approximation $F(x_k) \approx J'(x_k)J(x_k)$ and applies Eqn. 14 only to find a search direction, that is,

$$J(x_k)'J(x_k)p_k = -\nabla f(x_k)' = -J(x_k)'h(x_k). \tag{15}$$

The next point is found by a line search, $x_{k+1} = x_k + \alpha^* p_k$,

$$\alpha^* = \arg\min_{\alpha>0} f\left(x_k + \alpha p_k\right). \tag{16}$$

(b) The function is already written in this form:

$$f\left(x\right) = \left[10\left(x_2 - x_1^2\right)\right]^2 + \left[1 - x_1\right]^2, \tag{17}$$

and

$$\begin{aligned} h_1\left(x\right) &= 10\left(x_2 - x_1^2\right), \\ h_2\left(x\right) &= 1 - x_1. \end{aligned}$$

Then

$$J = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \\ \frac{\partial h_1}{\partial x_2} & \frac{\partial h_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} -20x_1 & 10 \\ -1 & 0 \end{bmatrix}, \tag{18}$$

which for $x = 0$ leads to

$$J\left(0\right)' J\left(0\right) = \begin{bmatrix} 0 & -1 \\ 10 & 0 \end{bmatrix} \begin{bmatrix} 0 & 10 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix}. \tag{19}$$

Thus, the search direction p is found from

$$J\left(0\right)' J\left(0\right) p = -J\left(0\right)' h\left(0\right), \tag{20}$$

or

$$\begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = -\begin{bmatrix} 0 & -1 \\ 10 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{21}$$

The solution is

$$p = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{22}$$

Note that p in this particular case is proportional to the negative gradient of $f\left(x\right)$ at $x = 0$, since $\nabla f\left(0\right) = \left(-2, 0\right)$. The line search consists of finding the minimum of

$$f\left(\alpha p\right) = 100\alpha^4 + \left(1 - \alpha\right)^2.$$

It turns out that

$$\frac{\partial\left(100\alpha^4 + \left(1 - \alpha\right)^2\right)}{\partial\alpha} = 400\alpha^3 - 2 + 2\alpha = 0$$

only for $\alpha \approx 0.161\,26$, which determines x_1,

$$x_1 \approx \begin{bmatrix} 0.161\,26 \\ 0 \end{bmatrix}. \tag{23}$$

Problem 3:

(a) Show that the KKT equations for the LP-problem

$$\min_x c'x \quad Ax \geq b,\ x \geq 0, \tag{24}$$

may be written as

$$\begin{aligned} Ax \geq b,\ x \geq 0, \\ A'\pi \leq c,\ \pi \geq 0, \\ \pi'(Ax - b) = 0, \\ (c - A'\pi)'x = 0. \end{aligned} \tag{25}$$

(b) State the Duality Theorem for LP and prove that the dual problem to the problem in (a) is the problem

$$\max_\pi b'\pi \quad A'\pi \leq c,\ \pi \geq 0. \tag{26}$$

(c) Determine the minimum value of the objective function

$$f(x) = 2x_1 + 3x_2 + 3x_3 + x_4 + 2x_5 \tag{27}$$

when

$$\begin{aligned} x_1 + 3x_2 + x_3 + x_4 + 0 \geq 4 \\ 2x_1 + x_2 + x_3 - x_4 + x_5 \geq 1, \\ x_i \geq 0 \end{aligned} \tag{28}$$

Hint: Use a graphical solution.

Solution:

(a) The equations first need to be stated in the basic form, that is,

$$\min_x c'x \tag{29}$$

$$Ax - b \geq 0, \tag{30}$$

$$x \geq 0. \tag{31}$$

Let us use the Lagrange multipliers π for (30) and s for (31). Then $\mathcal{L} = c'x - \pi'(Ax - b) - s'x$, and the KKT equations are

$$\begin{aligned} (\nabla_x \mathcal{L}_A)' = c - A'\pi - s = 0, \\ \pi'(Ax - b) = 0 \\ s'x = 0, \\ \pi, s \geq 0 \end{aligned} \tag{32}$$

Since $s = c - A'\pi$, this may be simplified to

$$\begin{aligned} c - A'\pi &\geq 0, \\ \pi' (Ax - b) &= 0, \\ \left(c - A'\pi\right)' x &= 0, \\ \pi &\geq 0. \end{aligned} \tag{33}$$

Together with (30) and (31) this is identical to Eqns. 25.

(b) The Duality theorem for LP says that to each (primal) LP problem there is a dual problem with equivalent KKT equations. In the dual problem the variables are the Lagrange multipliers of the primal problem, and the primal variables the Lagrange multipliers. The dual of the dual problem is equivalent to the primal problem. Dual objective values are always below the primal objective values and the optimal objective values for the primal and dual problems are equal.

Then consider the problem in Eqns. 26:

$$\min_{\pi} \left\{-b'\pi\right\} \tag{34}$$

$$c - A'\pi \geq 0, \tag{35}$$

$$\pi \geq 0. \tag{36}$$

with Lagrangian $\mathcal{L}_B = -b'\pi - x'(c - A'\pi) - t'\pi$ and KKT equations,

$$\begin{aligned} (\nabla_\pi \mathcal{L}_B)' &= -b + Ax - t = 0, \\ x' \left(c - A'\pi\right) &= 0, \\ t'\pi &= 0, \\ x, t &\geq 0. \end{aligned} \tag{37}$$

Similarly as above, since $t = Ax - b$,

$$\begin{aligned} \left(c - A'\pi\right)' x &= 0, \\ \pi' (Ax - b) &= 0, \\ x &\geq 0. \end{aligned} \tag{38}$$

Together with the inequalities (35) and (36), this is the same as Eqns. 25.

(c) From the context it should be obvious that we must use the dual problem. Moreover, only the optimal objective value is asked for, – and not x^*! The problem is already in the form of Eqn. 24, and the dual problem has only two variables, (π_1, π_2):

$$\begin{aligned} &\max_{\pi} b'\pi \\ &A'\pi \leq c, \ \pi \geq 0. \end{aligned} \tag{39}$$

Then,

$$\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \leq \begin{bmatrix} 2 \\ 3 \\ 3 \\ 1 \\ 2 \end{bmatrix}, \ b = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \ \pi_1, \pi_2 \geq 0. \tag{40}$$

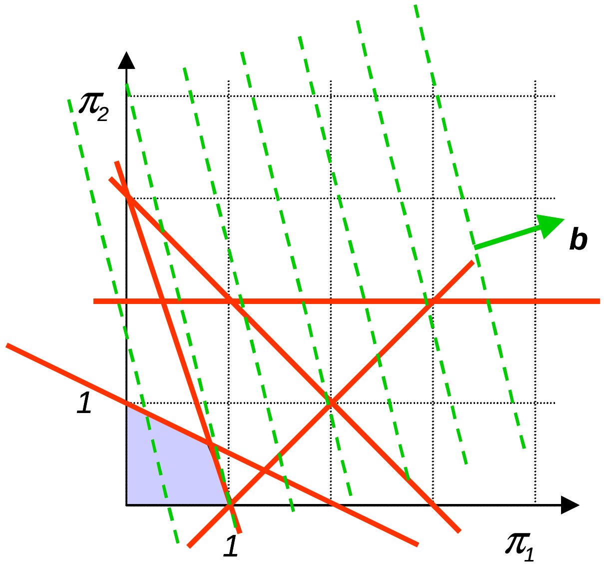

Figure 2: Constraints, final feasible domain, and contours and gradient for the objective.

We follow the hint and make a drawing, Fig. 2. Since we are looking for a *maximum*, this is clearly $\pi_1 = 1$ and $\pi_2 = 0$. Thus,

$$\min_x c'x = \max_\pi b'\pi = 4 \times 1 + 1 \times 0 = 4.$$

Problem 4:

(a) *Consider the standard functional,*

$$F(y) = \int_a^b f(t, y(t), \dot{y}(t))\, dt,\ \ \dot{y} = dy/dt \tag{41}$$

defined for the set $\mathcal{D} = \left\{ y \in C^1[a,b]\ ;\ y(a) \text{ and } y(b) \text{ fixed} \right\}$. *Explain what it means that* f *is strongly convex, and prove that this implies that* F *is strictly convex.*

A ship has a fuel consumption per time unit proportional to $C(v, \dot{v}) = (a\dot{v} + v)^2$, *where* v *is the velocity,* $\dot{v}$ *the acceleration and* a *is a constant* ($\neq 0$). *The total fuel consumption from* $t = 0$ *to* $t = T$ *is therefore*

$$H(v) = \int_0^T C(v(t), \dot{v}(t))\, dt. \tag{42}$$

(b) *Prove* H *is strictly convex on the set* $\mathcal{D} = \left\{ v \in \mathcal{C}^1[0,T]\ ; v(0) = 0 \right\}$ *even if* C *is not strongly convex.*

(c) *The ship is supposed to cover a distance* L *in time* T *using a minimal amount of fuel (starting with* $v(0) = 0$*). Formulate the corresponding optimization problem. If a solution exists, will it be unique?*

(d) *Assume for simplicity that* $a = 1$ *and* $L = 1$. *Verify that the solution to the problem in (c) in* *this case may be written*

$$v(t) = A\left[\left(e^{t}-1\right)+\left(2e^{T}-1\right)\left(e^{-t}-1\right)\right],$$
$$A = \left(3e^{T}-2Te^{T}+e^{-T}-4\right)^{-1}. \tag{43}$$

Solution:

(a) The definition of *strongly convex* is as follows: $f(x,y,z)$ is strongly convex for all x in an interval $[a,b]$ if

$$f(x,y+v,z+w)-f(x,y,z) \geq f_y(x,y,z)\,v+f_z(x,y,z)\,w, \tag{44}$$

for all y and z, with equality if and only if v or w are 0. To be precise, one needs to assume that f, f_y and f_z are continuous and also limit (x,y,z) and $(x,y+v,z+w)$ to some set $S \in \mathbb{R}^3$ (Troutman Def. 3.4). This rather curious condition turns out to be sufficient for F to be strictly convex. Assume that f is strongly convex for $x \in [a,b]$:

$$f\left(x,y+v,y'+v'\right)-f(x,y,z) \geq f_y\left(x,y,y'\right)v+f_z\left(x,y,y'\right)v' \tag{45}$$

By integrating both sides between a and b we obtain

$$F(y+v)-F(y) \geq \delta F(y;v), \tag{46}$$

so that F is certainly convex. However, the inequalities in (45), and hence also (46) are strict unless v or $\dot{v}$ are 0. If v or $\dot{v}$ are 0, then $v(x)\,v'(x)=0$ for all $x \in [a,b]$, such that $\frac{d}{dx}v^2(x)=0$, or $v(x)=$ constant. Since we have $v(a)=v(b)=0$, this implies that $v=0$, or exactly what is needed for F to be strictly convex.

(b) The case here shows that strong convexity of f is sufficient but not necessary for F to be strictly convex. First of all,

$$\begin{aligned}
C(v+\alpha,w+\beta)-C(v,w) &= \left(a(v+\alpha)+(w+\beta)\right)^2-(av+w)^2 \\
&= \left((av+w)+(a\alpha+\beta)\right)^2-(av+w)^2 \\
&= 2(av+w)(a\alpha+\beta)+(a\alpha+\beta)^2 \\
&= C_v\alpha+C_w\beta+(a\alpha+\beta)^2 \\
&\geq C_v\alpha+C_w\beta.
\end{aligned} \tag{47}$$

However, we will have equality for all α and β such that $\beta=-a\alpha$, showing that C is *not* strongly convex.

If we now consider the identity above with $w=\dot{v}$ and $\beta=\dot{\alpha}$, then

$$H(v+\alpha)-H(v) = \delta H(v;\alpha)+\int_0^T \left(a\alpha(t)+\dot{\alpha}(t)\right)^2 dt \tag{48}$$

The last term will be strictly positive unless $a\alpha(t)+\dot{\alpha}(t)=0$ for $t \in [0,T]$. But this implies that $\alpha(t)=Ae^{\cdot -at}$, and since $\alpha(0)=0$, A is also equal to 0 , and $\alpha(t)=0$ for $t \in [0\ T]$. Hence, H is strictly positive.

(c) The problem may be formulated as

$$\min_{v \in \mathcal{D}} H(v) \tag{49}$$

when

$$G(v) = \int_0^T v(t)\,dt = L \tag{50}$$

For the set $\mathcal{D}$ we must have that $v(0) = 0$. At the other end there is no fixed condition, and in order to ensure that the boundary terms after the partial integration vanishes, we need a "natural condition" at $t = T$, in the present case,

$$\frac{\partial C(v, \dot{v})}{\partial \dot{v}} = 2\left[v(T) + \dot{v}(T)\right] = 0. \tag{51}$$

Thus,

$$\mathcal{D} = \left\{ v \in \mathcal{C}^1[0, T] \ ;\ v(0) = 0, v(T) + \dot{v}(T) = 0 \ \right\}. \tag{52}$$

For the solution, we minimize the extended Lagrangian $\mathcal{L}(v, \lambda) = H(v) + \lambda G(v)$. Here, H is strictly convex on $\mathcal{D}$, whereas G is linear and hence convex regardless the sign on λ. Thus, $\mathcal{L}$ is strictly convex. If we are at all able to find a solution for some $\lambda \in \mathbb{R}$, it will be unique.

(d) According to (c), the extended kernel is $\tilde{f} = (v + \dot{v})^2 + \lambda v$, and the Euler equation becomes

$$\frac{d}{dt}\tilde{f}_{\dot{v}} - \tilde{f}_v = \frac{d}{dt}2(v + \dot{v}) - 2(v + \dot{v}) - \lambda = 0, \tag{53}$$

or

$$\ddot{v} - v = \frac{\lambda}{2}. \tag{54}$$

From (c), the boundary conditions are $v(0) = 0$ and $v(T) + \dot{v}(T) = 0$. The general solution to the Euler equation is easily seen to be

$$v = Ae^t + Be^{-t} - \frac{\lambda}{2}. \tag{55}$$

There are three constants to fit, and it is convenient to start with λ, which follows $v(0) = 0$:

$$A + B - \frac{\lambda}{2} = 0. \tag{56}$$

Thus,

$$v(t) = A\left(e^t - 1\right) + B\left(e^{-t} - 1\right). \tag{57}$$

The condition at $t = T$ leads to

$$\begin{aligned} v(T) + \dot{v}(T) &= A\left(e^T - 1\right) + B\left(e^{-T} - 1\right) + Ae^T - Be^{-T} \\ &= A\left(2e^T - 1\right) - B = 0, \end{aligned} \tag{58}$$

or, $B = A\left(2e^T - 1\right)$. Thus, we obtain the promising

$$v(t) = A\left[\left(e^t - 1\right) + \left(2e^T - 1\right)\left(e^{-t} - 1\right)\right]. \tag{59}$$

Finally, the value of A follows from the constraint:

$$\begin{aligned} 1 &= \int_0^T v(t)\,dt = A\int_0^T \left[\left(e^t - 1\right) + \left(2e^T - 1\right)\left(e^{-t} - 1\right)\right] dt \\ &= A\left[3e^T - 2Te^T + e^{-T} - 4\right]. \end{aligned} \tag{60}$$

Note that a only a verification of the solution was asked for. Thus, one only needs to check that the solution in Eqn. 43 satisfies $v(0) = 0$, $v(T) + \dot{v}(T) = 0$, and $\int_0^T v(t)\,dt = 1$.

We also read from the expression for $v(t)$ that

$$\frac{\lambda}{2} = \frac{2e^T}{3e^T - 2Te^T + e^{-T} - 4}. \tag{61}$$

Figure 3 shows 3 cases of $v(t)$ for different values of T. Note that L is equal to 1 for each case.

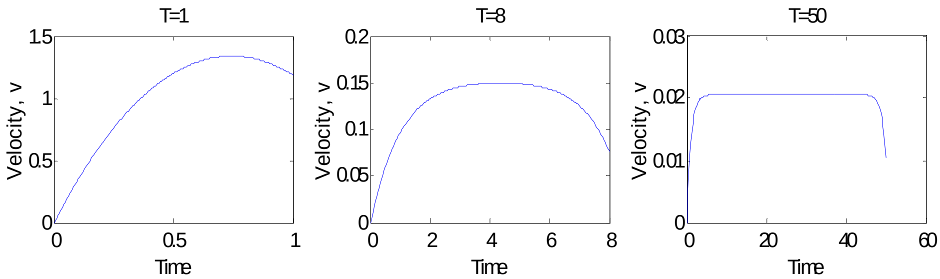

Figure 3: Velocity vs. time for various sizes of T. The distance has been scaled to 1 in each case.