6.832: Problem Set #2

Due on Wednesday, February 28, 2018 at 17:00. See course website for submission details. Use Drake release tag `drake-20180220` , i.e. use this notebook via `./docker_run_notebook.sh drake-20180220 .` , or whichever script you need for your platform.

About this problem set

This problem set will entirely live inside this jupyter notebook.

Grades will be assigned based on three components:

- **Manually graded free-response questions** -- the TAs will manually assign grades to your answers to short answer responses. You can write inline responses using Markdown (https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet) with inline LaTeX -- double-click on any problem writeup to see some examples. Double-click response areas to edit them, and press Control-Enter to finish editing them.
- **Automated code testing** -- we will run automated tests against specific functions (see more details when we introduce the first coding test).
- **Quick code review** -- we will perform a quick manual check to make sure you have actually implemented the functions correctly (as opposed to hacked the unit tests to pass!).

The automated coding tests are pretty small in this problem set, but we are planning to move more towards this framework in the next three problem sets. We would love to hear feedback from you on how this testing setup works.

1. Cost Functions

In this problem we will explore how to design cost functions that make the robot exhibit the kind of behavior we want. For this, we will consider the Dubins car model, which is a very simple model of a vehicle given by the following equations:

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ \psi \end{bmatrix}, \qquad \dot{\mathbf{x}} = f(\mathbf{x}, u) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} -\sin\psi \\ \cos\psi \\ u \end{bmatrix},$$

where $\mathbf{x}$ is the state of the system and consists of the states x (the x-position), y (the y-position) and ψ (the yaw angle of the vehicle).

The only control input is the steering angle u, which directly controls $\dot{\psi}$, while the car drives at constant velocity. (At $\psi = 0$, the vehicle drives in the $+y$ direction.) u has limits $[-u_{max}, +u_{max}]$.

Note that this is a very simple model -- it does not follow Newtonian physics, and can instead just instantaneously choose its yaw rate. It is however a simple model that has seen some use for actual robot research, for both UAVs and cars. For example, here's some relevant results on Google Scholar. (https://scholar.google.com/scholar?hl=en&as_sdt=0%2C22&q=dubins+path+planning&btnG=) To give us a little visual, here's a figure from a paper that uses a Dubins model for a 2D UAV in an environment with obstacles Barry et al., 2012 (http://groups.csail.mit.edu/robotics-center/public_papers/Barry12.pdf):

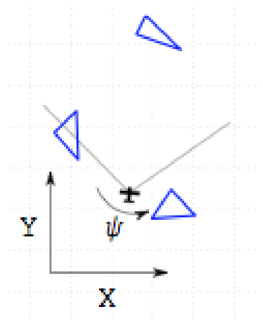


Figure: A dubins vehicle, navigating amonst polygonal obstacles.

Question 1.1 (2 points)

In general it is useful to have cost functions that penalize both the states along a trajectory, and also the final state:

$$J = \mathbf{x}(t_f)^T Q_f \mathbf{x}(t_f) + \int_0^{t_f} g(\mathbf{x}(t), \mathbf{u}(t))\, dt$$

In this question let's look only at cost functions that only involve the final state. A simple form for a cost function that penalizes only the final state of the robot is

$$J = \mathbf{x}(t_f)^T Q_f \mathbf{x}(t_f)$$

where t_f is the final time and Q_f is a symmetric positive semidefinite matrix of the appropriate size. Suppose we want the robot to end up with its yaw angle close to 0, but do not care about the final x and y positions. What should we choose Q_f to be (remember to make sure it is symmetric and positive semidefinite)?

In [1]:
```
'''
Autograded answer for 1.1

Please implement this function so that the return argument
satisfies the specification above.

Scroll to the end of the notebook to execute all tests for this notebo
ok.
'''
import numpy as np

def get_Q_f_problem_1_1():
    Q_f = np.zeros((3, 3))
    # YOUR CODE HERE TO SET Qf
    Q_f[2][2]=5
    return Q_f

print get_Q_f_problem_1_1()
```

```
[[ 0.  0.  0.]
 [ 0.  0.  0.]
 [ 0.  0.  5.]]
```

Short answer explanation for 1.1. Provide a brief justification of your choice of Q_f in this cell.

Since the final state does not care what x,y position the car ends in, then the cost function cannot favor or penalize any values for x,y. Let $Q_f = [[a_1, b_1, c_1], [a_2, b_2, c_2], [, a_3, b_3, c_3]]$. To achieve no penalty for any x/y positions, we need the cost to have no terms with x or y:

$$J = \mathbf{x}(t_f)^T Q_f \mathbf{x}(t_f) = x^2 a_1 + xya_2 + \psi x a_3 + xyb_1 + y^2 b_2 + \psi y b_3 + \psi x c_1 + y\psi c_2 + \psi^2 c_3$$

This means that even entry in Q_f must be 0 except for c_3 since this is the constant for the only term that is independent of x and y. Lastly, this entry is non-zero because we want to penalize yaw angles that are further from zero, so therefore the cost is $J = c_3\psi^2 = 5\psi^2$.

Furthermore, $c_3 >= 0$ to be positive semidefinite. This matrix Q_f is positive semidefinite becauses it is symmetric and its eigenvalues are greater than or equal to zero: $\lambda = 0, 5 \geq 0$.

Question 1.2 (2 points)

Now suppose we want the vehicle to end up close to the line $y = 0.5x$, but we do not care exactly where on this line and what yaw angle it ends up in. What should we choose Q_f to be (remember to make sure it is symmetric and positive semidefinite)?

In [2]:
```
'''
Autograded answer for 1.2

Please implement this function so that the return argument
satisfies the specification above.
'''

def get_Q_f_problem_1_2():
    Qf = np.zeros((3, 3))
    # YOUR CODE HERE TO SET Qf
    Qf[0][0]=1
    Qf[1][1]=4
    Qf[0][1]=-2
    Qf[1][0]=-2
    return Qf

print get_Q_f_problem_1_2()
```

```
[[ 1. -2.  0.]
 [-2.  4.  0.]
 [ 0.  0.  0.]]
```

Short answer explanation for 1.2. Provide a brief justification of your choice of Q_f in this cell.

Let $Q_f = [[a_1, b_1, c_1], [a_2, b_2, c_2], [, a_3, b_3, c_3]]$. Since we do not care about the final yaw angle, then column c_i will all be zeros such that the final cost function will not be affected by whatever the final yaw rate is. To maintain symmetry, this means $a_3, b_3 = 0$ also. Then, since we want the final state to be on the line $y = .5x$, then we want points on the line, such as $(2, 1)$ to have cost of zero. With the remaining non-zero constants (and the fact that it must be symmetric so $a_2 = b_1$), the cost function is:

$$J = a_1x^2 + a_2xy + b_1xy + b_2y^2 = a_1x^2 + 2a_2xy + b_2y^2$$

So when we evaluate the cost function at a final state $(x, y) = (2, 1)$, we have $4a_1 + 4a_2 + b_2 = 0$, and the values in my above Q_f satisfy this constraint and also keep the matrix positive semidefinite since it is symmetric and the eigenvalues $\lambda = 0, 5 \geq 0$.

Question 1.3 (3 points)

Now suppose we want to end up close to the curve $y = x^2$, and again do not care about the final yaw angle or where exactly on this curve we end up. Why is it not possible to set Q_f to achieve this?

Short answer explanation for 1.3. Please put your answer to this question in this cell.

Since we still don't care about the yaw angle, the cost function is still $J = a_1x^2 + a_2xy + b_1xy + b_2y^2$. And while we can find constants a_1, a_2, b_1, b_2 such that a single point on the line $y = x^2$ yields zero cost, there is no way to make it so that the cost is zero for any point on the line since there is no term with just y (and not y^2).

2. Optimal Control via HJB

Consider the scalar equation

$$\dot{x} = -4x + 2u,$$

and the infinite horizon cost function

$$J = \int_0^\infty [32x^2 + u^2]dt.$$

Question 2.1 (5 points)

Assume that the optimal cost-to-go function is of the form $J^\star = px^2$. What value of p satisfies the Hamilton-Jacobi-Bellman conditions for optimality?

Written answer explanation for 2.1. A proper answer to this question involves writing out more than a couple expressions. We recommend working through this problem on pencil and paper first. Please then write your answer below in this cell, using LaTeX-style math expressions.

The HJB conditions says that for $\dot{x} = f(x,u)$ and additive cost $\int_0^\infty g(x,u)dt$, then the cost-to-go function J will be optimal when $0 = min_u[g(x,u) + \frac{\partial J^*}{\partial x} f(x,u)]$. The partial derivative of $J^* = px^2$ with respect to x is $\frac{\partial J^*}{\partial x} = 2px$, and then $f(x,u) = -4x + 2u$ and $g(x,u) = 32x^2 + u^2$. If we plug these into the main minimization expressionof HJB:

$$min_u[g(x,u) + \frac{\partial J^*}{\partial x} f(x,u)] = min_u[32x^2 + u^2 + 2px(-4x + 2u)] = min_u[32x^2 + u^2 - 8px^2 + 4$$

Then, to get the optimal policy, which is the u^* that will minimize this expression, we need to take the partial derivative with respect to u:

$$\frac{\partial}{\partial u}(32x^2 + u^2 - 8px^2 + 4pxu) = 2u + 4px$$

Therefore we get the optimal control policy by setting this equal to zero, so that will be $u^* = -2px$. And the HJB sufficiency theorem tells us that the optimal cost-to-go occurs when u is optimal, so we can plug this control policy back into the original expression and solve for a value of p to get the optimal cost-to-go:

$$\begin{gathered} 0 = 32x^2 + u^2 - 8px^2 + 4pxu \\ 0 = 32x^2 + (-2px)^2 - 8px^2 + 4px(-2px) \\ 0 = 32 + 4p^2 - 8p - 8p^2 \\ p = 2 \end{gathered}$$

Therefore, we can write our cost to go function as $J = px^2 = 2x^2$.

Question 2.2 (3 points)

Given that the optimal feedback controller associated with $J^\star$ is $u^\star = -Kx$, what is the value of K?

Written answer explanation for 2.2. We recommend working through this problem on pencil and paper first. Please then write your answer below in this cell, using LaTeX-style math expressions.

Since in the above question we determined that the optimal control policy is $u = -2px$ and we now know that $p = 2$, then we get:

$$K = 4$$

3. Typing in the Optimal Controller for the Double Integrator

In this problem, we'll consider the optimal control problem for the 1-dimensional input-constrained double integrator described by

$$\ddot{q} = u, \ |u| \leq 1$$

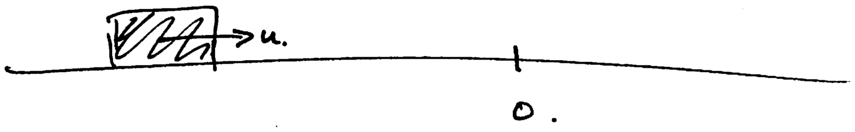


Figure: Simple drawing of a double integrator

As you may remember, a simple way to think of the double integrator is as a system which has direct control over its acceleration, like a brick sliding around on ice, with a rocket booster attached for control. While a double integrator (especially in just 1 dimension) is a very simple system, double or triple integrators (which control jerk, the derivative of acceleration) can be for example used as simple models (see Fig. 2) to approximate the dynamics of quadrotors (http://groups.csail.mit.edu/robotics-center/public_papers/Florence16.pdf)) for performing obstacle avoidance and planning paths (https://www.youtube.com/watch?v=9a0eEscz1Cs&t=1s).

We won't often be able to simply type in an optimal controller, but for the double integrator, it's not too hard!

For this question, we want the optimal controller for a minimum-time cost function. Remember that for "minimum-time" problems, we just have the cost function:

$$J = \int_0^\infty g(x(t), u(t)) \, dt$$

for just:

$$g(x,u) = \begin{cases} 0, & \text{if } x = 0 \\ 1, & \text{otherwise} \end{cases}$$

Which gives us:

$$J = t_f$$

Where t_f is the time needed to reach the desired state, 0. So our cost function is just the time needed to reach 0.

Question 3.1 (4 points)

Implement the optimal controller, and the optimal cost-to-go function, for the double integrator in the space below. (May help to look in the course notes at Example 8.2)

The point of this question isn't to practice our ability to implement algorithms, but instead to give us some data and experience to later make some interesting comparisons and analyses.

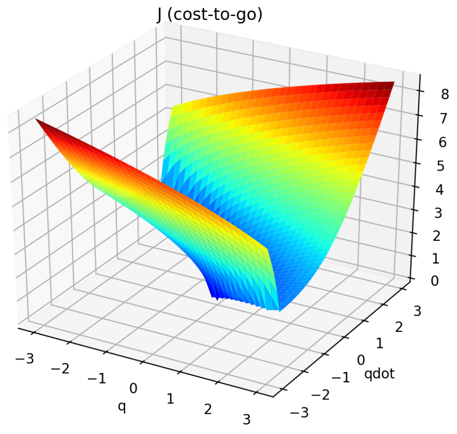
Also you can rest well knowing that no-one will ever have a "more optimal" controller than you! (But actually, this maybe isn't the best controller. We'll discuss why.)

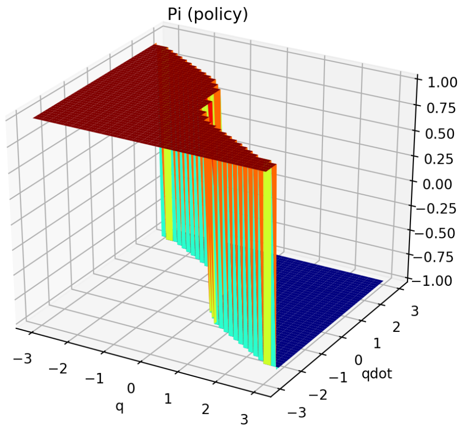
In [3]:
```
'''
Autograded answer for 3.1

Please implement this function so that the return argument
satisfies the specification above.
'''

def get_optimal_time_to_go_problem_3_1(q, qdot):
    if (qdot<0 and q<=.5*qdot**2) or (qdot>=0 and q < -.5*qdot**2):
        return 2*(.5*qdot**2-q)**(1/2.0)-qdot
    elif (q==0 and qdot==0):
        return 0
    else:
        return 2*(.5*qdot**2+q)**(1/2.0)+qdot

def get_optimal_control_problem_3_1(q, qdot):
    #u = q**3+np.cos(qdot) # bogus answer to make plots pretty
    if (qdot<0 and q<=.5*qdot**2) or (qdot>=0 and q<-.5*qdot**2):
        return 1
    elif (q==0 and qdot==0):
        return 0
    else:
        return -1
```

Plotting help for Question 3.1

We've written a little plotting code for you to visualize this policy:

In [4]:
```
%matplotlib notebook
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from matplotlib import cm
import matplotlib.animation as animation

num_q_bins = 31
qbins = np.linspace(-3., 3., num_q_bins)
num_qdot_bins = 51
qdotbins = np.linspace(-3., 3., num_qdot_bins)
state_grid = [set(qbins), set(qdotbins)]

[Q, Qdot] = np.meshgrid(qbins, qdotbins)

fig_optimal, (ax_J, ax_Pi) = plt.subplots(1, 2, figsize=(12,5))
ax_J.axis('off')
ax_Pi.axis('off')

ax_J = fig_optimal.add_subplot(121, projection='3d')
ax_J.set_title('J (cost-to-go)')
ax_J.set_xlabel("q")
ax_J.set_ylabel("qdot")

ax_Pi = fig_optimal.add_subplot(122, projection='3d')
ax_Pi.set_title('Pi (policy)')
ax_Pi.set_xlabel("q")
ax_Pi.set_ylabel("qdot")
plt.tight_layout()
global_list = []

J_optimal = np.zeros((num_qdot_bins, num_q_bins))
for i, row in enumerate(J_optimal):
    for j, val in enumerate(row):
        J_optimal[i,j] = get_optimal_time_to_go_problem_3_1(Q[i,j], Qd
ot[i,j])

J_surface = [ax_J.plot_surface(Q, Qdot, J_optimal, rstride=1, cstride=
1,
                       cmap=cm.jet)]

Pi_optimal = np.zeros((num_qdot_bins, num_q_bins))
Pi_optimal = np.zeros((num_qdot_bins, num_q_bins))
for i, row in enumerate(Pi_optimal):
    for j, val in enumerate(row):
        Pi_optimal[i,j] = get_optimal_control_problem_3_1(Q[i,j], Qdot
[i,j])

Pi_surface = [ax_Pi.plot_surface(Q, Qdot, Pi_optimal, rstride=1, cstri
de=1,
                        cmap=cm.jet)]
```





Note that the plotting code, for both J and Pi, is just based on a 3D mesh over a `np.meshgrid` of Q and Qdot.

We recommend printing out these inputs to the plots, and practice manipulating the data of the plots.

In [5]:
```
# Some examples of interacting with the data generated above
print qbins.shape
print qdotbins.shape
print Q.shape
print Qdot.shape
print J_optimal.shape
print Pi_optimal.shape

terrible_control_choice = -Pi_optimal
qbin_sample = 10
qdotbin_sample = 3
print terrible_control_choice[qdotbin_sample][qbin_sample]
print get_optimal_control_problem_3_1(5,15)
print get_optimal_time_to_go_problem_3_1(5,15)

q = qbins[qbin_sample]
qdot = qdotbins[qdotbin_sample]
q=2
qdot = 1
#print np.linspace(-3,3,32)
#print "Loop"
#x = np.linspace(-10,10, 2000)
#for i in x:
#    print "control: "+str(get_optimal_control_problem_3_1(2,i))+ " fo
r index "+str(i)
#print "Optimal Time"
#print get_optimal_time_to_go_problem_3_1(q,qdot)
#print "Optimal Control according to value iteration"
#print get_optimal_control_problem_3_1(q,qdot)
#print get_optimal_control_problem_3_1(2,-2)

# Recommend investigating more the shapes of some of these and how the
y relate
```

```
(31,)
(51,)
(51, 31)
(51, 31)
(51, 31)
(51, 31)
-1.0
-1
36.6794833887
```

Question 3.2 (1 point)

For a real robot, why would the minimum-time solution not necessarily be the "best" thing to do?

Written answer explanation for 3.2.

Minimum time may not be the best solution to do if you are trying to optimize for different metrics, such as accuracy (and not over-shooting your target) for a task, user-experience or comfort for interactive robots and self driving cars, or minimizing energy and gas consumption throughout the course of the set of actions. Minimum time is ideal if all that you care about is completing the task as fast as possible.

4. Value Iteration solution for the Double Integrator

Now let's approach the same problem, but solve it using the Value Iteration algorithm.

An implementation of that algorithm is available for you in Drake. This is a complete implementation of the algorithm with discrete actions and volumetric interpolation over state.

In [6]:
```
import math
import numpy as np


from pydrake.systems.framework import VectorSystem
from pydrake.systems.analysis import Simulator
from pydrake.systems.controllers import (
    DynamicProgrammingOptions, FittedValueIteration)

# This system block implements the dynamics
# of a double integrator. It takes one control
# input, has two states (1D position and velocity),
# and copies its current state as its output.
class DoubleIntegrator(VectorSystem):
    def __init__(self):
        # One input, one output, two state variables.
        VectorSystem.__init__(self, 1, 1)
        self._DeclareContinuousState(2)

    # qddot(t) = u(t)
    def _DoCalcVectorTimeDerivatives(self, context, u, x, xdot):
        xdot[0] = x[1]
        xdot[1] = u

    # y(t) = x(t)
    def _DoCalcVectorOutput(self, context, u, x, y):
        y[:] = x

# Set up a simulation of this system.
plant = DoubleIntegrator()
simulator = Simulator(plant)
options = DynamicProgrammingOptions()

# This function evaluates a minimum time
# running cost, given a context (which contains
# information about the current system state).
def min_time_cost(context):
    x = context.get_continuous_state_vector().CopyToVector()
    if x.dot(x) < .05:
        return 0.
    return 1.

# This function evaluates a running cost
# that penalizes distance from the origin,
# as well as control effort.
def quadratic_regulator_cost(context):
    x = context.get_continuous_state_vector().CopyToVector()
    u = plant.EvalVectorInput(context, 0).CopyToVector()
    return 2*x.dot(x) + 10*u.dot(u)

# Pick your cost here...
cost_function = min_time_cost
#cost_function = quadratic_regulator_cost

# This sets up the mesh of sample points over
# which we'll run the value iteration algorithm.
```

```
num_q_bins = 31 #32
qbins = np.linspace(-3., 3., num_q_bins)
num_qdot_bins = 51 #34
qdotbins = np.linspace(-3., 3., num_qdot_bins)
state_grid = [set(qbins), set(qdotbins)]
#print qbins
#print qdotbins

input_limit = 1.
input_grid = [set(np.linspace(-input_limit, input_limit, 9))]
timestep = 0.01

[Q, Qdot] = np.meshgrid(qbins, qdotbins)

# Recommend not increasing the max_iterations too much.  Another facto
r of 10 should be okay, if you want to wait.
max_iterations = 10000
J_iterations  = np.zeros((num_qdot_bins, num_q_bins, max_iterations))
Pi_iterations = np.zeros((num_qdot_bins, num_q_bins, max_iterations))
num_iterations = 0
def add_iteration_solve(iteration, mesh, cost_to_go, policy):
    global num_iterations, J_iterations, Pi_iterations
    num_iterations += 1
    if num_iterations >= max_iterations:
        raise RuntimeError("Solution did not converge within "+str(max
_iterations)+" iterations.")

    J = np.reshape(cost_to_go, Q.shape)
    J_iterations[:,:,iteration-1] = J    # "first" iteration here is
 "1", so here we 0-order it
    Pi = np.reshape(policy, Q.shape)
    Pi_iterations[:,:,iteration-1] = Pi


options.visualization_callback = add_iteration_solve

# Run value iteration on this mesh using
# the cost function chosen above.
policy, cost_to_go = FittedValueIteration(simulator, cost_function,
                                          state_grid, input_grid,
                                          timestep, options)

# trim array of iterations
J_iterations  = J_iterations[:,:,:num_iterations]
Pi_iterations = Pi_iterations[:,:,:num_iterations]
#print J_iterations[25][15][9990:]


print "Done solving.  Converged in "+str(num_iterations)+" iteration
s."
```

```
Done solving.  Converged in 1067 iterations.
```

Similar to before, it's useful to manipulate the data used to generate the plots.

Unlike with the analytical optimal policy, where we only had one mesh over q and $\dot{q}$, this time we have saved the solution state from each solve. Let's print out the shape of `Pi_iterations` . Here we've set it up so that indexing into the last dimension gives us the solver at that iteration's state.

In [7]:
```
print Pi_iterations.shape
print Q.shape
print Q[8][25]



policy_on_10th_iteration = Pi_iterations[:,:,10]

final_computed_policy    = Pi_iterations[:,:,-1]

print J_iterations[8][25][-1]
print Pi_iterations[8][25][-1]
print Pi_iterations[10][25][-1]
```

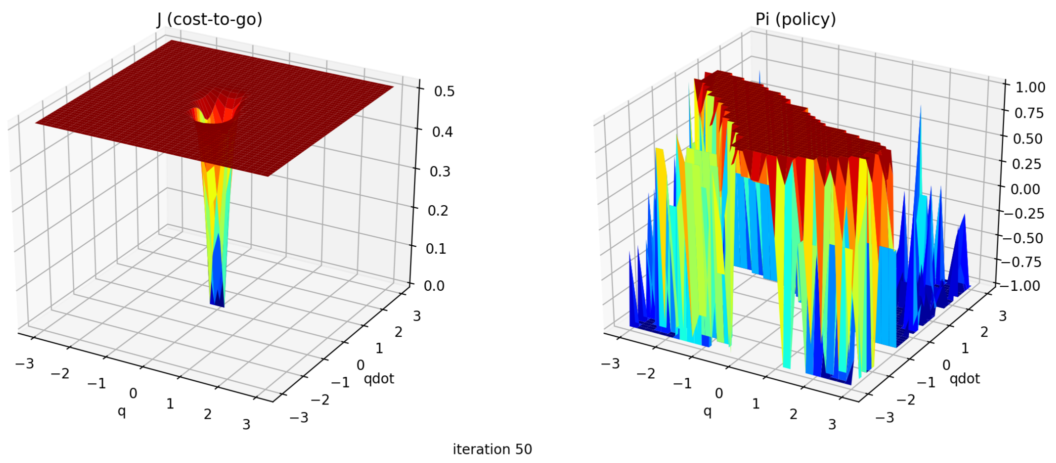
```
(51, 31, 1067)
(51, 31)
2.0
3.28871855476
1.0
1.0
```

Now plot the result of Value Iteration

The above code cell already solved the value iteration code, and saved its iterative solution on each iteration.

Now below, we plot this solution over time.

In [8]:
```
%matplotlib notebook
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from matplotlib import cm
import matplotlib.animation as animation

fig, (ax, ax2) = plt.subplots(1, 2, figsize=(12,5))
ax.axis('off')
ax2.axis('off')

ax = fig.add_subplot(121, projection='3d')
ax.set_title('J (cost-to-go)')
ax.set_xlabel("q")
ax.set_ylabel("qdot")

ax2 = fig.add_subplot(122, projection='3d')
ax2.set_title('Pi (policy)')
ax2.set_xlabel("q")
ax2.set_ylabel("qdot")
plt.tight_layout()

J_initial = J_iterations[:,:,0]
J_surface = [ax.plot_surface(Q, Qdot, J_initial, rstride=1, cstride=1,
                       cmap=cm.jet)]

Pi_initial = Pi_iterations[:,:,0]
Pi_surface = [ax2.plot_surface(Q, Qdot, Pi_initial, rstride=1, cstride
=1,
                        cmap=cm.jet)]

def update_plots(iteration, J_iterations, J_surface, Pi_iterations, Pi
_surface):
    if iteration % 10 != 0:
        return # only plot every 10th
    J_surface[0].remove()
    J_surface[0] = ax.plot_surface(Q, Qdot, J_iterations[:,:,iteration
], rstride=1, cstride=1,
                        cmap=cm.jet)

    for txt in fig.texts:
        txt.set_visible(False)
    fig.text(0.5, 0.04, 'iteration '+str(iteration), ha='center', va=
'center')

    Pi_surface[0].remove()
    Pi_surface[0] = ax2.plot_surface(Q, Qdot, Pi_iterations[:,:,iterat
ion], rstride=1, cstride=1,
                        cmap=cm.jet)


animate = animation.FuncAnimation(fig, update_plots, num_iterations, i
nterval=1, fargs=(J_iterations, J_surface,

      Pi_iterations, Pi_surface))
```



Tips

- while the animation is running, you can click + drag to change the 3D viewpoint
- pressing the "Power" button for the plot above will stop the animation, which is probably making your computer run pretty hard if you have it running
- restarting your jupyter notebook's kernel (the restart circle button up top in jupyter notebook) will also completely free up your system (but also lose all variable state)

Run the value iteration code for the double integrator to compute the optimal policy and optimal cost-to-go for the minimum-time problem.

Compare the result to the analytical solution we found in lecture, and programmed in Question 3.1 by answering the following questions.

Question 4.1 (5 points)

Find an initial condition of the form $(2, \dot{q}_0)$ such that the value iteration policy takes an action in exactly the wrong direction from the true optimal policy.

a) What value of $\dot{q}_0$ did you find that disagrees?

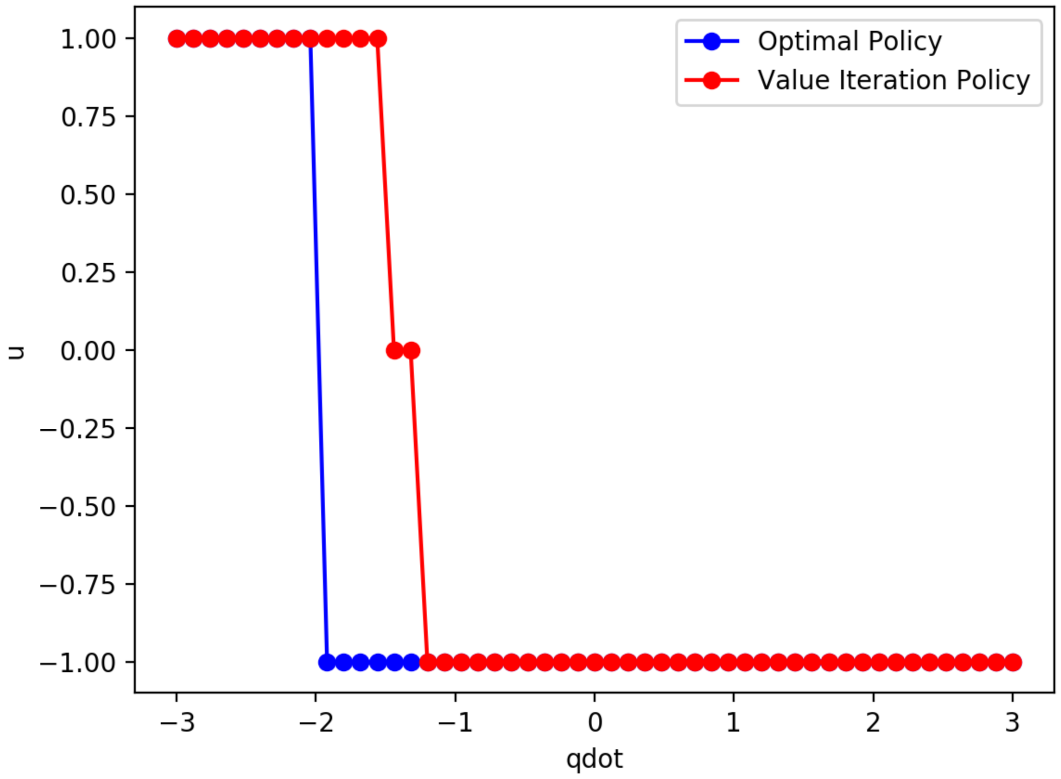
b) What is the true optimal time-to-go from this state (i.e., for the optimal bang-bang controller derived in class)?

c) What is the time-to-go estimated by value iteration?

d) Why are these values different?

In [9]:
```
import matplotlib
import numpy as np
import math
import matplotlib.pyplot as plt
%matplotlib notebook


policy_on_10th_iteration = Pi_iterations[:,:,10]

final_computed_policy    = Pi_iterations[:,:,-1]

#All tuple pairs for which q=2 (given that q=np.linspace(-3,3,31) and
 qdot=np.linspace(-3,3,51))
q_two=[(0, 25), (1, 25), (2, 25), (3, 25), (4, 25), (5, 25), (6, 25),
(7, 25), (8, 25), (9, 25), (10, 25), (11, 25), (12, 25), (13, 25), (14
, 25), (15, 25), (16, 25), (17, 25), (18, 25), (19, 25), (20, 25), (21
, 25), (22, 25), (23, 25), (24, 25), (25, 25), (26, 25), (27, 25), (28
, 25), (29, 25), (30, 25), (31, 25), (32, 25), (33, 25), (34, 25), (35
, 25), (36, 25), (37, 25), (38, 25), (39, 25), (40, 25), (41, 25), (42
, 25), (43, 25), (44, 25), (45, 25), (46, 25), (47, 25), (48, 25), (49
, 25), (50, 25)]

#Plotting the different qdot values and then the outputted policies
x_list = []
expected = []
valueiter = []
for tup in q_two:
    if Qdot[tup[0],tup[1]] == -1.8:
        print tup[0]
        print tup[1]
    if Q[tup[0],tup[1]]==2:
        #print "value iter: "+str(Pi_iterations[tup[0]][tup[1]][-1])
        #print "optimal: "+str(get_optimal_control_problem_3_1(2,Qdot
[tup[0],tup[1]]))
        #print "Q: "+str(Q[tup[0],tup[1]])
        #print "Qdot: "+str(Qdot[tup[0],tup[1]])
        x_list.append(Qdot[tup[0],tup[1]])
        expected.append(get_optimal_control_problem_3_1(2,Qdot[tup[0],
tup[1]]))
        valueiter.append(Pi_iterations[tup[0]][tup[1]][-1])
plt.plot(x_list,expected,"bo-",label="Optimal Policy")
plt.plot(x_list,valueiter,"ro-",label="Value Iteration Policy")
plt.xlabel("qdot")
plt.ylabel("u")
plt.legend()
```

```
10
25

Traceback (most recent call last):
  File "/usr/local/lib/python2.7/dist-packages/matplotlib/cbook/__init
__.py", line 388, in process
    proxy(*args, **kwargs)
  File "/usr/local/lib/python2.7/dist-packages/matplotlib/cbook/__init
__.py", line 228, in __call__
    return mtd(*args, **kwargs)
  File "/usr/local/lib/python2.7/dist-packages/matplotlib/animation.p
y", line 1498, in _stop
    self.event_source.remove_callback(self._loop_delay)
AttributeError: 'NoneType' object has no attribute 'remove_callback'
```



Out[9]:
```
<matplotlib.legend.Legend at 0x7f014450f690>
```

In [10]:
```
#Value iterations estimated final time (cost function is the time-to-g
o)
print J_iterations[10][25][-1]
```

```
2.85650375959
```

Written answer explanation for 4.1. Bonus points if you generate plots that help explain your answer. (Feel free to add code cells, in addition to this markdown cell.)

a) For an initial state of $(q, \dot{q}) = (2, -1.8)$, the value iteration algorithm chooses a control policy of $u = +1$.However, the optimal policy would be to break commanding a $u = -1$ because for a starting state $q = 2$, then optimally $u = 1$ when $\dot{q} \leq -2$ and $u = -1$ when $\dot{q} > -2$. This can be seen in the above plot showing the policies outputted by the optimal policy function and the value iteration function for varying inputs of $\dot{q}$ when $q = 2$ is fixed.

b) The optimal time-to-go is $t = \dot{q} + 2\sqrt{.5\dot{q}^2 + q}$ for when $u = -1$ is the control policy. Therefore, a starting state of $(q, \dot{q}) = (2, -1.8)$ has an optimal time-to-go of $t = -1.8 + 2\sqrt{.5(-1.8)^2 + 2} = 2.00525951809$.

c) The time-to-go estimated by value iteration is represented by J_iterations, and for $(q, \dot{q}) = (2, -1.8)$, it is estimated as 2.8565037, as shown in the above code block.

d) These values are likely very different because the value iteration initially decides for this starting state that the best policy is exactly opposite that of the optimal policy. Therefore, the time to get to the goal will take more time because it initial control goes in the opposite direction of what the policy should be. That is why likely why the value iteration time-to-go estimate is much larger than the optimal time-to-go calculation because the set of actions required to get to the goal that value iteration finds are not the minimum/optimal set of actions. Additionally, the value iteration cost size is susceptible to errors caused by the given discretization, and therefore can be very different from the optimal time because this is found exactly.

Question 4.2 (3 points)

When implementing value iteration, one needs to be wary of several implementation details.

Find a setting of the discretization (i.e., the variables `num_q_bins` and `num_qdot_bins`) that causes the code to NOT converge in 10,000 iterations.

The [underlying implementation (https://github.com/RobotLocomotion/drake/blob/master/systems/controllers/dynamic_programming.cc)](https://github.com/RobotLocomotion/drake/blob/master/systems/controllers/dynamic_programming.cc) of value iteration terminates after performing an update to J *(following the discrete value iteration update described in class) and finding that the update did not change J* by more than $\epsilon = 0.0001$ at any point.

Why doesn't the setting you discovered converge?

Note: The maximum distance between points in the q and $\dot{q}$ directions should still be at most 0.2, and the grid must still contain the square with sides of length 2 centered about the origin.

Short answer for 4.2.

When num_q_bins=32 and num_qdot_bins=34, the value iteration algorithm does not converge. This is because the mesh grid size, when doing the linspace between $[-3, 3]$, does not contain the final goal state of $(q, \dot{q}) = (0, 0)$ due the discretization. Therefore the cost to go, which represents the time to get to this final state, will continue to diverge forever since it is unattainable in this implementation.

In [11]:
```
q = np.linspace(-3,3,32)
print q
qdot = np.linspace(-3,3,34)
print qdot
```

```
[-3.         -2.80645161 -2.61290323 -2.41935484 -2.22580645 -2.032258
06
 -1.83870968 -1.64516129 -1.4516129  -1.25806452 -1.06451613 -0.870967
74
 -0.67741935 -0.48387097 -0.29032258 -0.09677419  0.09677419  0.290322
58
  0.48387097  0.67741935  0.87096774  1.06451613  1.25806452  1.451612
9
  1.64516129  1.83870968  2.03225806  2.22580645  2.41935484  2.612903
23
  2.80645161  3.        ]
[-3.         -2.81818182 -2.63636364 -2.45454545 -2.27272727 -2.090909
09
 -1.90909091 -1.72727273 -1.54545455 -1.36363636 -1.18181818 -1.
 -0.81818182 -0.63636364 -0.45454545 -0.27272727 -0.09090909  0.090909
09
  0.27272727  0.45454545  0.63636364  0.81818182  1.          1.181818
18
  1.36363636  1.54545455  1.72727273  1.90909091  2.09090909  2.272727
27
  2.45454545  2.63636364  2.81818182  3.        ]
```

Question 4.3 (2 points)

We may have noticed some issues for the value iteration solution, compared to what we know is the optimal solution. But it's worth taking a moment to consider the pros and cons of each method we've tried so far:

a) What is at least one reason that the value iteration algorithm is powerful, compared to the analytical solution?

b) On the other hand, what is nice about the analytical solution?

Short answers for 4.3.

a) The value iteration algorithm can converge to a solution that is within a constant factor away from the optimal solution very rapidly and without requiring you to solve complex optimization problems and differential equations to compute the exact analytical policy.

b) The analytical solution will be exact and optimal for every feasible inputted state presuming you successfully solved for the optimal policies.

Question 4.4 (2 points)

Switch the cost function for the double integrator so that we use a quadratic cost on both: control input, and state. This has already been implemented for you as: `quadratic_regulator_cost()` .

How does the minimum-time solution of value iteration compare to the quadratic-regulator solution?

Short answer for 4.4.

The policy computed for the minimum time cost function is a piecewise quadratic function, whereas for the quadratic regular appears to have a more linear divide (negative slope) between u=+/-1 for the policy. As well, the iterations to ultimately get to the final converged policy solution are much noiser and osciallting between each step of the algorithm for the minimum time function. Whereas for the quadratic regulator cost function, it smoothly converges to the policy over each iteration.

The cost-to-go J for the quadratic regulator is much more accurate even in early iterations than the minimum time cost function, which initially reflects extremes of either being very little time if the starting state is approximately the goal or being maximum time if the state is far from the goal.

Lastly, for the same input and grid setup for q and $\dot{q}$, the quadratic regulator takes longer(more iterations) to converge to the ultimate solution

Note: you can still receive full credit if you do not do the bonus questions. But you may be able to get extra credit if you do them.

5. Bonus Question: Value Iteration for the simple pendulum

Bonus 5.1 (2 points)

See if you run the value iteration example for the input-constrained pendulum. Either replicate the setup above we provided for the double integrator, or use the `bash` scripts provided for the docker image, i.e. run:

```
# or use the bash script for your system
./docker_run_bash_linux.sh drake-20180220 .
# now inside docker image
cd /underactuated/src
cd pendulum
python value_iteration.py
```

Question: how does the value-iteration solution compare to other control design methods we have explored for the input-constrained pendulum?

Short answer for 5.1.

YOUR ANSWER HERE

Test your own implementations

Running the cell below will run your implemented functions against unit tests.

Don't change the cell below, or the `test_set_2.py` file. We will grade your implementations against the original files.

Make sure to **SAVE** your notebook before running tests. (File --> Save and Checkpoint, or use the hotkey which should be ctrl+s on linux, cmd+s on osx, etc)

In [12]:
```
import os
# Run the test in a subprocess, to make sure it doesn't open any plot
s...
os.popen("python test_set_2.py set_2.ipynb test_results.json")

# Print the results json for review
import test_set_2
print test_set_2.pretty_format_json_results("test_results.json")
```

```
Test Problem 1_1: 2.00/2.00.

Test Problem 1_2: 2.00/2.00.

Test Problem 3_1 Optimal Control Input: 2.00/2.00.

Test Problem 3_1 Time (Cost) To Go: 2.00/2.00.

TOTAL SCORE (automated tests only): 8.00/8.00
```

Note that many of the questions are not auto-graded, but will be graded manually! (Double-check you gave answers for each.)

In [ ]:
```

```