MIT MANAGEMENT SLOAN SCHOOL

15.455x Mathematical Methods of Quantitative Finance

Week 8: Optimization

Paul F. Mende
MIT Sloan School of Management

Finance at MIT
Where ingenuity drives results

Critical points

Finance at MIT
Where ingenuity drives results

Taylor's theorem

Recall that for a function of a single variable, Taylor's theorem

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2!}(x - x_0)^2 f''(x_0) + \frac{1}{3!}(x - x_0)^3 f'''(x_0) + \cdots$$

means that near a **critical point** where the first derivative vanishes,

$$f(x) - f(x_0) \approx \frac{1}{2} f''(x_0)(x - x_0)^2, \quad \text{provided } f''(x_0) \neq 0$$

In higher dimensions, this approximation generalizes so that the neighborhood of a critical point is described by a **quadratic form**.

Taylor's theorem

- For a function of several variables,

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\nabla f)^\top (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top Q(\mathbf{x} - \mathbf{x}_0) + \cdots ,$$

$$\text{where} \quad \nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \end{pmatrix}, \qquad Q = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots \\ \vdots & & \ddots \end{pmatrix} = Q^\top$$

so that

$$f(\mathbf{x}) - f(\mathbf{x}_0) \approx \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^\top Q(\mathbf{x} - \mathbf{x}_0)$$

where the matrix Q of second derivatives is a symmetric matrix.

Critical points: eigenvalues determine type

- If the **eigenvalues** of Q are all **positive**, the function is convex up and the critical point is a **minimum.**

- If the eigenvalues of Q are all **negative**, the function is concave and the critical point is a **maximum.**

- If Q has both positive and negative eigenvalues, then there are **saddle points**, which are both maximum and minimum, along different directions.

- If any eigenvalues are zero, there are **flat directions**.

- The **eigenvectors** determine the axes of orientation

Critical points

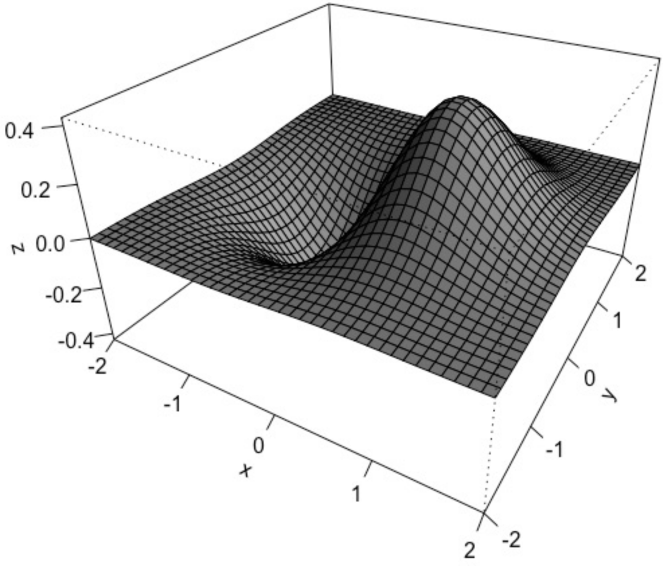
- Classification of critical points for continuous functions

```
f <- function(x,y) x*exp(-x^2-y^2)

x <- seq(-2,2, length=40)
y <- seq(-2,2, length=40)
z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30, expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')

filled.contour(x,y,z,  nlevels=9,
color=gray.colors)
```



$$f(x,y) = xe^{-(x^2+y^2)}$$

Critical points

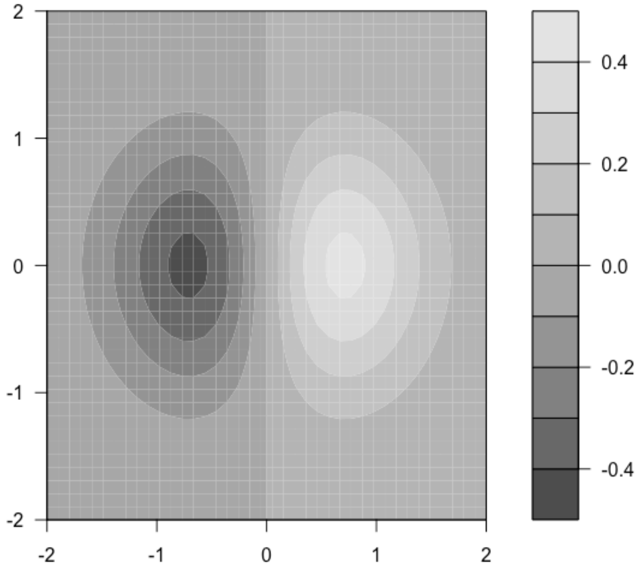
- Classification of critical points for continuous functions

```
f <- function(x,y) x*exp(-x^2-y^2)

x <- seq(-2,2, length=40)
y <- seq(-2,2, length=40)
z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30, expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')

filled.contour(x,y,z,  nlevels=9,
color=gray.colors)
```



$$f(x,y) = xe^{-(x^2+y^2)}$$

Critical points

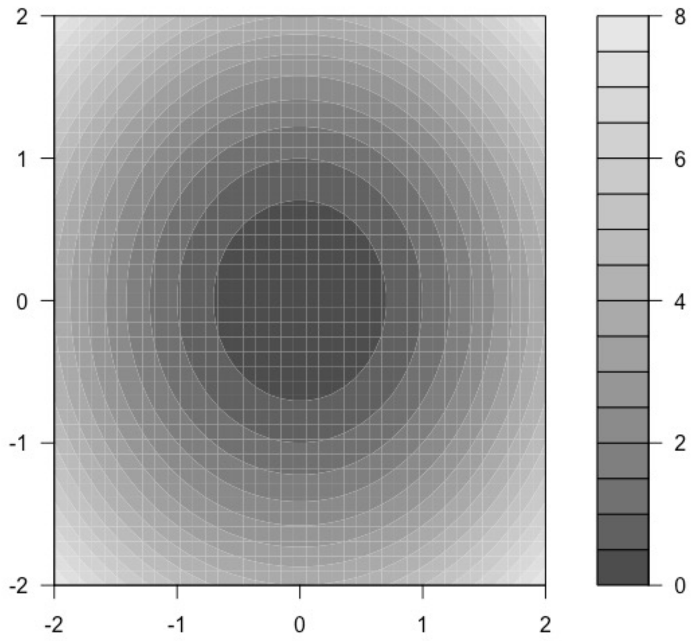
- Classification of critical points for continuous functions

```
f <- function(x,y) x^2 + y^2

z <- outer(x,y,f)

filled.contour(x,y,z,  nlevels=17,
    color=gray.colors)

persp(x,y,z,
    theta=30, phi=30, expand=0.6,
    col='gray', shade=0.75,
    ltheta=120,
    ticktype='detailed')
```



$$f(x,y) = x^2 + y^2$$

Critical points

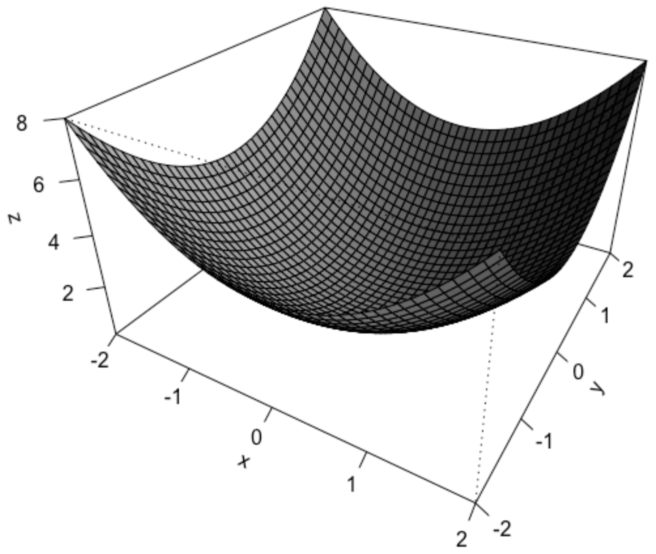
- Classification of critical points for continuous functions

```
f <- function(x,y) x^2 + y^2

z <- outer(x,y,f)

filled.contour(x,y,z,  nlevels=17,
color=gray.colors)

persp(x,y,z,
theta=30, phi=30, expand=0.6,
col='gray', shade=0.75,
ltheta=120,
ticktype='detailed')
```



$$f(x, y) = x^2 + y^2$$

Critical points: saddle point

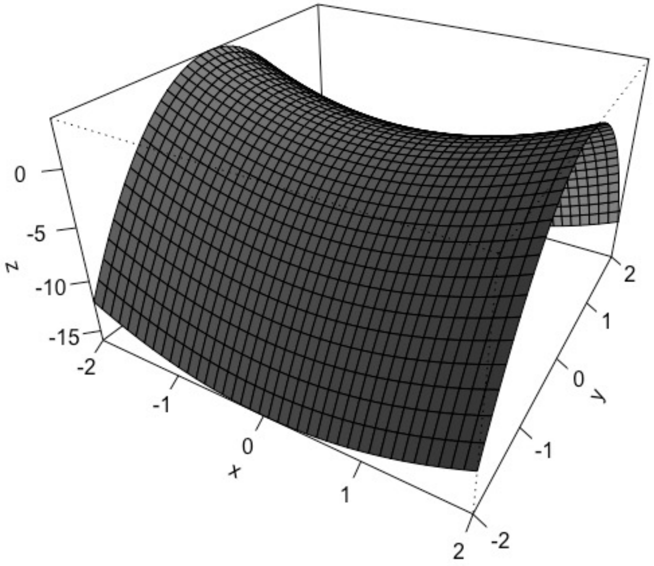
- Classification of critical points for continuous functions

```
f <- function(x,y) x^2 - 4* y^2

z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30, expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')

filled.contour(x,y,z,  nlevels=9,
color=gray.colors)
```



$$f(x,y) = x^2 - 4y^2$$

Critical points: saddle point

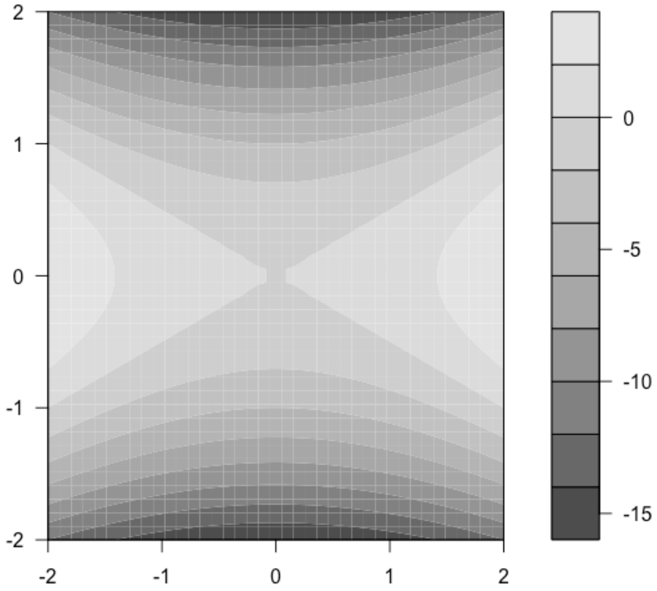
- Classification of critical points for continuous functions

```
f <- function(x,y) x^2 - 4* y^2

z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30, expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')

filled.contour(x,y,z,  nlevels=9,
color=gray.colors)
```



$$f(x, y) = x^2 - 4y^2$$

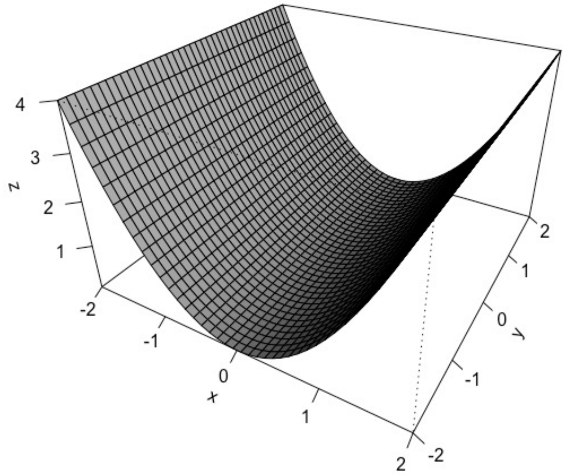
Critical points: flat directions

- Classification of critical points for continuous functions

```
f <- function(x,y) x^2

z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30,
expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')
```



$$f(x,y) = x^2$$

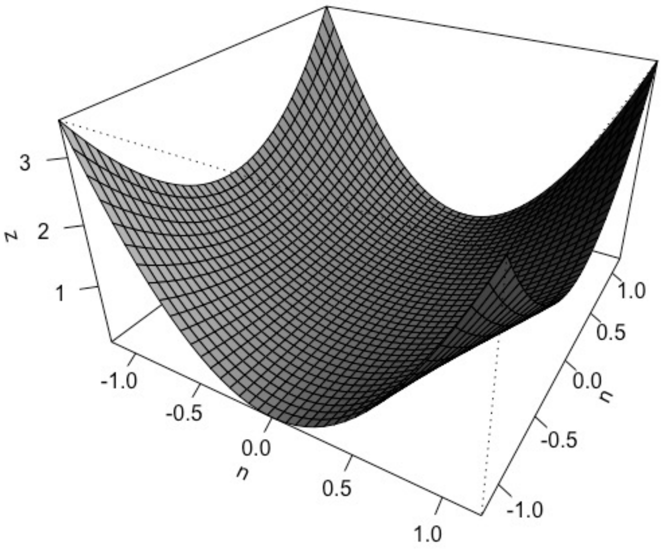
Critical points: symmetry breaking

- Classification of critical points for continuous functions

```
f <- function(x,y) x^2*(1+y^2)

z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30,
expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')
```



$$f(x,y) = x^2(1+y^2)$$

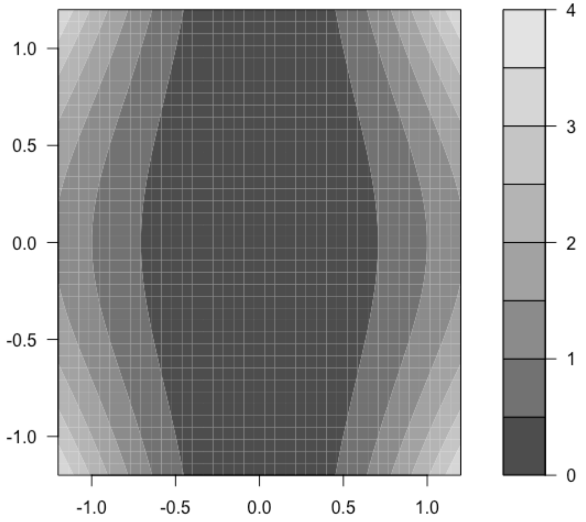
Critical points: symmetry breaking

- Classification of critical points for continuous functions

```
f <- function(x,y) x^2*(1+y^2)

z <- outer(x,y,f)

persp(x,y,z,
      theta=30, phi=30,
expand=0.6,
      col='gray', shade=0.75,
ltheta=120,
      ticktype='detailed')
```



$$f(x,y) = x^2(1+y^2)$$

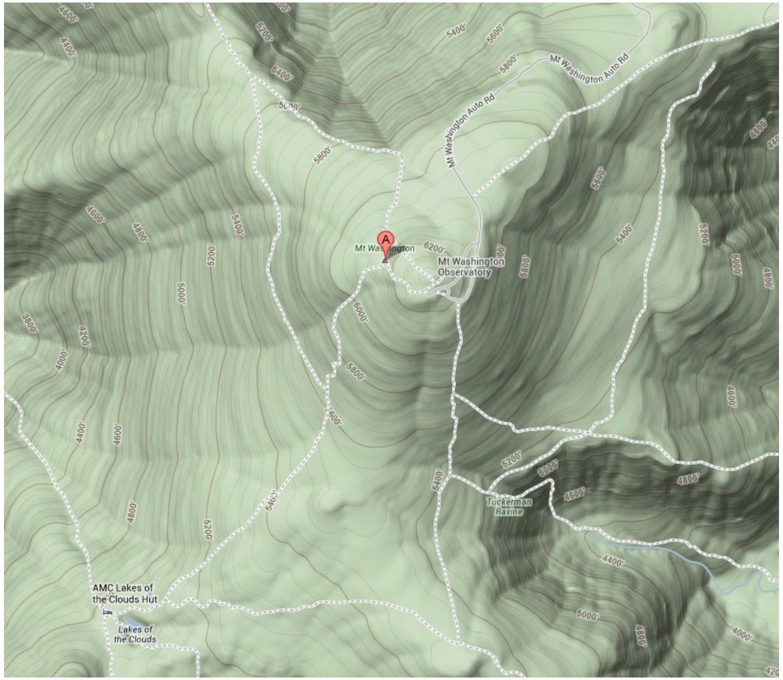
Constrained optimization and Lagrange multipliers

Finance at MIT
Where ingenuity drives results

Maximum along a path

Consider the height function *h(x,y)* around Mt. Washington

- **Contour lines** are curves of constant *h*.
- The maximum **along a path** occurs if the path is **tangent** to a contour line: the height will be stationary.



Lagrange multipliers

This observation leads to the method of Lagrange multipliers for solving **constrained optimization** problems.

- Let *h* and *g* be continuous, differentiable functions. Consider the extrema of *h(x,y),* subject to the **constraint** *g(x,y)* = *c*, where *c* is a constant.

- Define the **Lagrange function**

$$L(x,y,\lambda) \equiv h(x,y) - \lambda\big(g(x,y) - c\big)$$

- The extrema occur where **all the partial derivatives** of *L* vanish.

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} = \frac{\partial L}{\partial \lambda} = 0 \iff \nabla h \propto \nabla g$$

the direction of my constraint is along the direction of the level sets

Lagrange multiplier method

MANAGEMENT
SLOAN SCHOOL

- Introduce one constant, called a **Lagrange multiplier**, per constraint.

- Define Lagrange function, linear in constraints
  - Depends on **more variables** than original function
  - Has **simpler** solution (compared to alternatives like elimination)
  - Constraint function is not unique, need not be linear itself

- Find all critical points.

- Substitute and verify that they solve the problem. Often the **location** of an extremum is more interesting that its **value**.

Lagrange multipliers: example

$$
\begin{aligned}
h(x, y) &= x + y, \\
g(x, y) &= x^2 + y^2 = r^2, \\
L(x, y, \lambda) &= x + y - \lambda(x^2 + y^2 - r^2).
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial L}{\partial x} = 1 - 2\lambda x = 0 \implies x = \frac{1}{2\lambda}, \\
&\frac{\partial L}{\partial y} = 1 - 2\lambda y = 0 \implies y = \frac{1}{2\lambda} \implies y = x, \\
&\frac{\partial L}{\partial \lambda} = -x^2 - y^2 + r^2 = 0 \implies x = y = \pm\frac{r}{\sqrt{2}}, \\
&h_{\max} = \sqrt{2}r, \quad h_{\min} = -\sqrt{2}r
\end{aligned}
$$

Lagrange multipliers: example

$$
\begin{aligned}
h(x,y) &= 8x^2 + 12xy + 17y^2, \\
g(x,y) &= x^2 + y^2 = 1, \qquad\qquad \textit{(How about } g(x,y) = (x^2+y^2)^4\textit{?)} \\
L(x,y,\lambda) &= 8x^2 + 12xy + 17y^2 - \lambda(x^2+y^2-1).
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial x} &= 16x + 12y - 2\lambda x = 0, \\
\frac{\partial L}{\partial y} &= 12x + 34y - 2\lambda y = 0, \\
\frac{\partial L}{\partial \lambda} &= -(x^2 + y^2 - 1) = 0.
\end{aligned}
$$

$$
\begin{pmatrix} x \\ y \end{pmatrix} = \pm \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}, \quad \begin{pmatrix} x \\ y \end{pmatrix} = \pm \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}.
$$

$$
h_{\max} = 20, \quad h_{\min} = 5.
$$

Portfolio optimization

MIT MANAGEMENT SLOAN SCHOOL

- Mean variance optimization
- Start with expected asset mean returns and covariance matrix
  - (Where do these come from?)
  - (Are they well-estimated?)
  - (Log-normal? Robust? Stationary?)
- Constraints, e.g., budget, full-investment, beta, factor exposures.

- *Portfolio returns and variance*

$$\mu_p = \mu^\top \mathbf{w}$$

$$\sigma_p^2 = \mathbf{w}^\top C \mathbf{w}$$

- *Maximize return, minimize risk*

- *Budget constraint*

$$\sum_{i \in P} w_i = 1$$

Portfolio risk

Under the assumption of (log) normal returns,

$$\sigma_p^2 = \mathbf{w}^\top C \mathbf{w} = \sum w_i^2 \sigma_i^2 + 2 \sum_{i<j} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

The **covariance matrix** is

- symmetric
- positive-definite
  - excluding risk-free assets
  - excluding linearly dependent assets

Therefore it defines

- positive-definite **quadratic form** on *W*
- **inner product** on *W*: $\langle \mathbf{w}, \mathbf{w}' \rangle \equiv \mathbf{w}^\top C \mathbf{w}' = (C\mathbf{w})^\top \mathbf{w}'$
- norm on *W* identified with **risk**

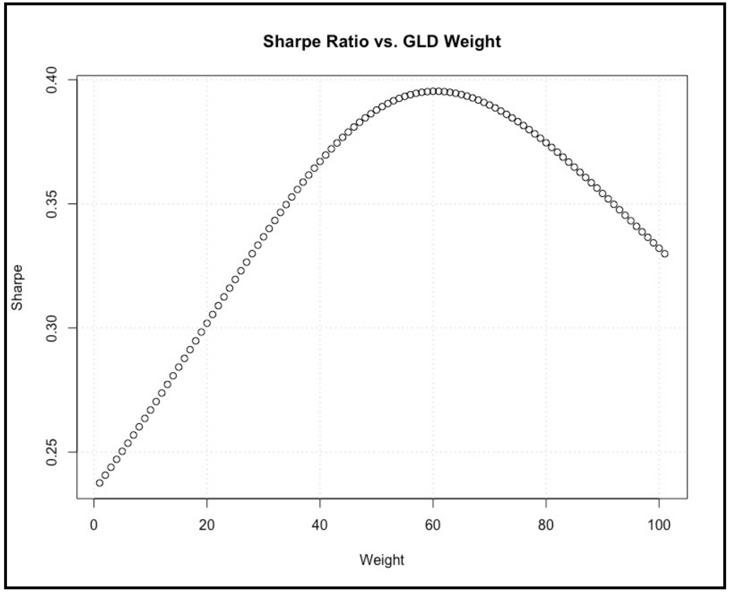
Example: two assets

- Estimation from historical data; construction; rebalancing

- Example:
  - Portfolio of SPX and GLD
  - Expected Sharpe ratio vs. past realized Sharpe ratio
  - Averaging time series implies rebalancing 50:50

```
port <- r$SPX
for (w in seq(.01,1,.01))
  port <- cbind(port,
  w*r$GLD + (1-w)*r$SPX)

plot(apply(port,2,mean)/apply(port,2,sd
)*sqrt(252))
```



Portfolio optimization – minimum variance

- Minimum-variance portfolio: All risk, no return

$$\mathcal{L}(\mathbf{w},\ell) = \frac{1}{2}\mathbf{w}^\top C\mathbf{w} + \ell(1-\iota^\top \mathbf{w}) \qquad \begin{cases} C & \text{covariance matrix} \\ \iota & \text{unit exposure vector,} \\ \ell & \text{Lagrange multiplier} \end{cases} \quad \iota = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix}$$

- Solution solves the constrained problem

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{\partial \mathcal{L}}{\partial \ell} = 0$$

Portfolio optimization - minimum variance

- Vary the weights:

$$\frac{\partial \mathcal{L}}{\partial w_i} = \sum_j C_{ij} w_j - \ell \iota_i = 0$$

- Solve for the weights by inverting matrix

$$\mathbf{w} = \ell C^{-1} \iota$$

- Solve for and eliminate Lagrange multiplier

$$\iota^\top \mathbf{w} = \ell \left(\iota^\top C^{-1} \iota \right) = 1$$

- Solution is $\mathbf{w}_{min} = \ell C^{-1} \iota = \dfrac{C^{-1}\iota}{\iota^\top C^{-1} \iota},$

assume no co–variance, weight on asset i is proportional to 1/sigma_i^2

notice that Lagrange part is eliminated

$$\sigma^2_{min} = \ell = 1/\left(\iota^\top C^{-1} \iota \right)$$

$$\mathbf{w}^\top C \mathbf{w}$$

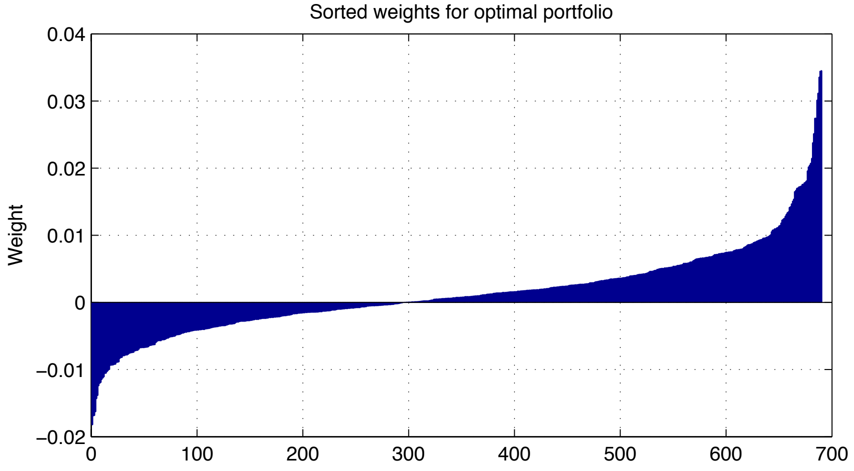
Portfolio optimization – minimum variance solution

- **Example:**
  - CRSP data set of daily returns for 690 US stocks
  - Estimate covariance matrix from historical data
  - Solve for weights of minimum variance portfolio

```
sigma_min <- sqrt(1/(t(Iota) %*% solve(Covariance) %*% Iota))

w_min = (solve(Covariance) %*% Iota) %*% sigma_min^2

sigma_min * sqrt(252) =    0.0436
```



Portfolio optimization - risk & return

- Now let's include return....How much return can we get? How little risk?
- What results are mathematically feasible? Are they achievable in practice?
- Given *any* portfolio, where does it sit relative to the efficient frontier?
- Generalize Lagrange function to

for a given level of return that we're demanding, what is the portfolio that has the least amount of risk

$$\mathcal{L}(\mathbf{w}, \ell, m) = \frac{1}{2}\mathbf{w}^\top C\mathbf{w} + \ell(1 - \iota^\top \mathbf{w}) + m(\mu_p - \mu^\top \mathbf{w})$$

Budget constraint **Return constraint**

Portfolio optimization - risk & return solution

- Vary the weights

$$\frac{\partial \mathcal{L}}{\partial w_i} = \sum_j C_{ij} w_j - \ell \iota_i - m \mu_i = 0$$

- Solve for the weights by inverting the covariance matrix

$$\mathbf{w} = C^{-1}(\ell \iota + m \mu)$$

- Solve for Lagrange multipliers by imposing constraints

$$\begin{aligned} \iota^\top \mathbf{w} &= \ell(\iota^\top C^{-1} \iota) + m(\mu^\top C^{-1} \iota) = 1 \\ \mu^\top \mathbf{w} &= \ell(\mu^\top C^{-1} \iota) + m(\mu^\top C^{-1} \mu) = \mu_p \end{aligned} \implies \begin{pmatrix} 1 \\ \mu_p \end{pmatrix} = M \begin{pmatrix} \ell \\ m \end{pmatrix}$$

Portfolio optimization - risk & return solution

- Solve for Lagrange multipliers by inverting 2x2 matrix

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

$$\begin{pmatrix} \ell \\ m \end{pmatrix} = M^{-1} \begin{pmatrix} 1 \\ \mu_p \end{pmatrix}$$

$$a \equiv \iota^\top C^{-1} \iota$$
$$b \equiv \mu^\top C^{-1} \iota$$
$$c \equiv \mu^\top C^{-1} \mu$$

- Eliminate Lagrange multipliers to obtain variance as a function of return:

$$\sigma_p^2 = \mathbf{w}^\top C \mathbf{w} = (\ell \quad m)\, M \begin{pmatrix} \ell \\ m \end{pmatrix}$$

We will show that $\mathbf{w}^\top C\mathbf{w} = (l \quad m)\, M \begin{pmatrix} l \\ m \end{pmatrix}$. By the definition of M, $M \begin{pmatrix} l \\ m \end{pmatrix} = \begin{pmatrix} 1 \\ \mu_p \end{pmatrix}$, so

$$(l \quad m)\, M \begin{pmatrix} l \\ m \end{pmatrix} = (l \quad m) \begin{pmatrix} 1 \\ \mu_p \end{pmatrix} = l + m\mu_p.$$

Because $\mathbf{w} = C^{-1}(l\iota + m\mu)$, we have

$$\mathbf{w}^\top C\mathbf{w} = \mathbf{w}^\top CC^{-1}(l\iota + m\mu) = \mathbf{w}(l\iota + m\mu) = l(\iota^\top \mathbf{w})^\top + m(\mu^\top \mathbf{w})^\top = l + m\mu_p,$$

where the last equality follows from the budget constraint and the return constraint. We can conclude that $\mathbf{w}^\top C\mathbf{w} = (l \quad m)\, M \begin{pmatrix} l \\ m \end{pmatrix}$.

Portfolio optimization – risk & return solution

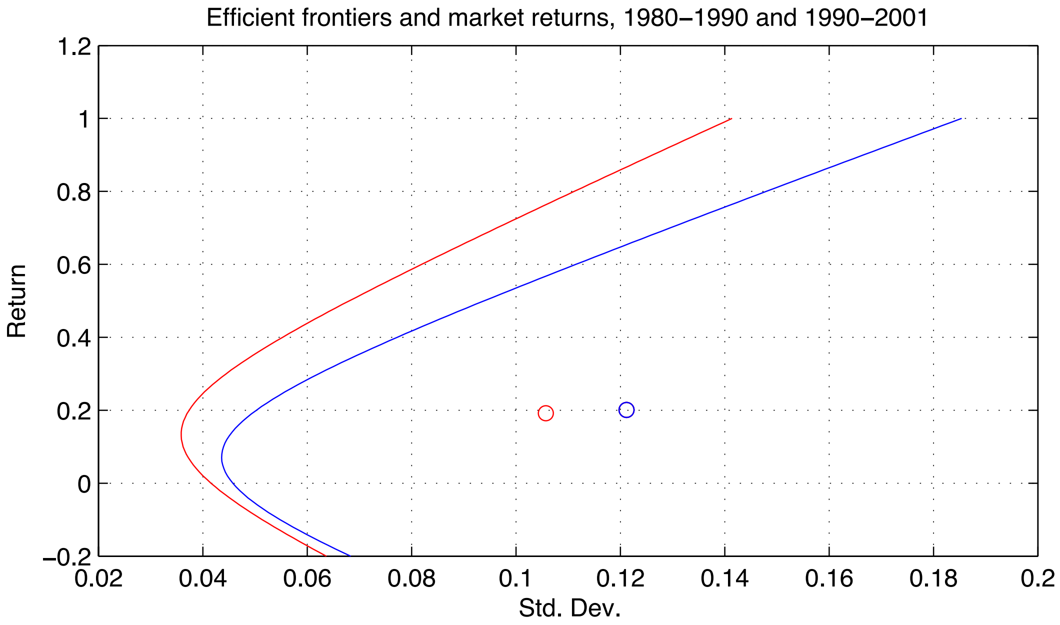
- Results of minimum-variance portfolio for a given return:

$$\sigma_p^2 = \left(\frac{1}{ac - b^2}\right)\left(a\mu_p^2 - 2b\mu_p + c\right)$$

  - Hyperbola in $\sigma_p - \mu_p$ space
  - Asymptotic behavior $\sigma_p \propto \mu_p$ mu_p --> infinity

Efficient frontier

- Boundary of feasible portfolios
  - Maps *N*-vectors of weights to 2-dimensional space
- Assumes normal returns, single-period horizons
- Changes over time
- Can always plot non-MVO portfolios in the space of feasible portfolios
- Sensitivity to inputs
- Stability



Characteristic portfolios and the efficient frontier

- Solution for efficient frontier, the set of fully-invested minimum-variance portfolios:

$$\mathbf{w} = C^{-1}(\ell\iota + m\mu)$$

- This is a **linear combination** of **two** characteristic portfolios, with weights given by the Lagrange multipliers as determined for a given value of expected return.
  - Global minimum variance portfolio
  - Portfolio of maximum Sharpe ratio

Portfolio optimization - with constraints

- If there are additional **linear constraints**, solution is easily generalized... Beta, factor loadings, industry exposure, etc. Add one Lagrange multiplier per constraint.

- If there are range **inequalities** on the weights themselves, then use **quadratic programming**.

- The solution space is a **subset** of unconstrained problem, so set of feasible portfolios lies **inside** unconstrained efficient frontier.

- Example: **long-only**, unlevered portfolio.
  - Minimize

$$\mathcal{L}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^\top C\mathbf{w}$$

  subject to

$$\sum_i w_i = 1, \qquad 0 \leq w_i \leq 1$$

Quadratic programming

MIT MANAGEMENT SLOAN SCHOOL

- Mean-variance optimization in practice is usually performed under **investment constraints**.
  - Position size (min, max)
  - Leverage
  - Factor exposure
  - Long/short neutrality
  - *VaR, Drawdown,...*

- *Many of these can be handled using **quadratic programming**.*

R: Solve a Quadratic Programming Problem   Find in Topic

solve.QP {quadprog}   R Documentation

Solve a Quadratic Programming Problem

Description

This routine implements the dual method of Goldfarb and Idnani (1982, 1983) for solving quadratic programming problems of the form *min(-d^T b + 1/2 b^T D b)* with the constraints *A^T b >= b_0*.

Usage

```
solve.QP(Dmat, dvec, Amat, bvec, meq=0, factorized=FALSE)
```

Arguments

`Dmat`	matrix appearing in the quadratic function to be minimized.
`dvec`	vector appearing in the quadratic function to be minimized.
`Amat`	matrix defining the constraints under which we want to minimize the quadratic function.
`bvec`	vector holding the values of *b_0* (defaults to zero).
`meq`	the first `meq` constraints are treated as equality constraints, all further as inequality constraints (defaults to 0).
`factorized`	logical flag: if `TRUE`, then we are passing *R^(-1)* (where *D = R^T R*) instead of the matrix *D* in the argument `Dmat`.

Value

a list with the following components:

`solution`   vector containing the solution of the quadratic programming

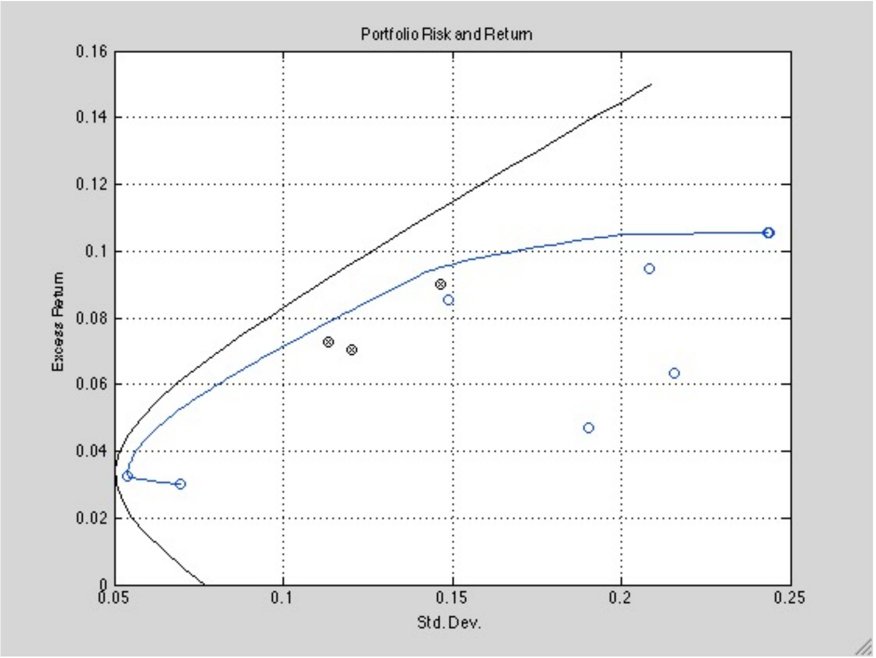
Case: Optimal asset allocation

- Country asset allocation
  - 8 assets - blue circles
  - Mean and vol of excess returns

- Unconstrained frontier
  - Black curve

- Constrained mean-variance efficient frontier
- `Blue line long–only constraint

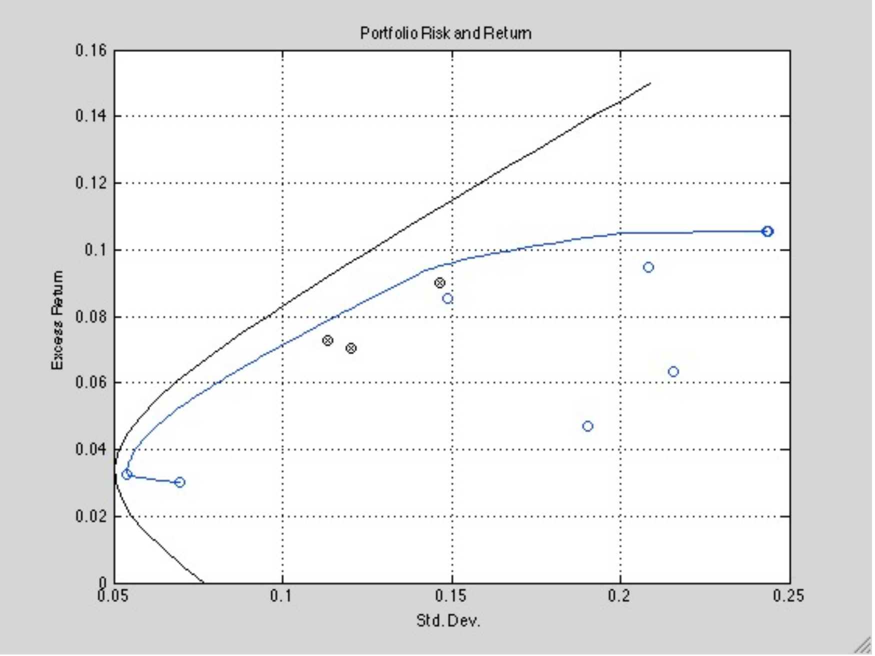
- Typical portfolios
  - Black filled circles
  - Equal, index, "current"



Case: Optimal asset allocation

Of note:

- Constrained frontier **strictly bounded** by min and max mean return of inputs; note endpoint portfolios w/ 100% allocation.
- Some non-optimized portfolios end up near the frontier (e.g., equal weighted) on their own. It is informative to **plot your portfolios** vs. the frontier -- even if they are not constructed with optimization.
- Optimizer is **extremely sensitive** to inputs -- and their errors. In this numerical example, two assets are nearly identical, yet endpoint goes "all in" rather than diversifying; note flatness.



References

- Books
  - Campbell, Lo, and MacKinlay (1997) – "*Econometrics of Financial Markets,*" Princeton
  - Fabozzi, Facardi, and Kolm (2006) – "*Financial Modeling of the Equity Market,*" Wiley
  - Grinold and Kahn (2000) – "*Active Portfolio Management,*" McGraw-Hill
  - Michaud (1998) – "*Efficient Asset Management,*" Harvard Business School Press