

## Multidimensional Optimization

Zsolt Pándi

Morgan Stanley Hungary Analytics Ltd.



This material has been prepared for information purposes to support the promotion or marketing of the transaction or matters addressed herein. It is not a solicitation of any offer to buy or sell any security, commodity or other financial instrument or to participate in any trading strategy. This is not a research report and was not prepared by the Morgan Stanley research department. It was prepared by Morgan Stanley sales, trading, banking or other non-research personnel. This material was not intended or written to be used, and it cannot be used by any taxpayer, for the purpose of avoiding penalties that may be imposed on the taxpayer under U.S. federal tax laws. Each taxpayer should seek advice based on the taxpayer's particular circumstances from an independent tax advisor. Past performance is not necessarily a guide to future performance. Please see additional important information and qualifications at the end of this material.

# Session Outline

- Problem Definition
- One-dimensional Optimization
- Multidimensional Optimization
  - Methods without Using Derivatives
  - Methods with Using Derivatives
  - Heuristics Using Random Numbers
- Tips&Tricks, Suggestions
- References



# Problem Definition

- We have a (possibly) non-linear, but deterministic function  $f(\mathbf{x})$  that depends on one or more independent variables.
- $f(\mathbf{x})$  is probably expensive to evaluate, and we may or may not be able to efficiently and/or accurately (!) compute its gradient or Hessian.
- Problem: find a minimum of  $f(\mathbf{x})$  with as few evaluations as possible.
- Examples of real-life problems
  - MLE, model fitting, cost minimization, performance maximization
- How do we recognize what we are looking for?
  - local vs. global optima
  - trying to find the longest straw in the haystack

# One-dimensional Optimization: Golden Section Search

- Analogy with bisection root search.
- Given an initial bracket of a minimum  $[a,b,c]$ , i.e.  $a < b < c$ , and  $f(b) < f(a)$ , and  $f(b) < f(c)$ , we take a guess at a point  $x$  either in  $(a,b)$  or in  $(b,c)$ . We evaluate  $f(x)$  and update our bracket accordingly.
  - e.g.  $x$  is in  $(a,b)$ : if  $f(x) < f(b)$  then the new bracket is  $[a,x,b]$ , else the new bracket is  $[x,b,c]$
- We repeat this until the bracket is small enough.
  - e.g. suppose that  $b$  is fraction  $w$  of the way between  $a$  and  $c$ , so  $w = (b-a)/(c-a)$ . Let the next trial point be a fraction  $z$  before  $b$ , so  $z = (b-x)/(c-a)$ . Hence the size of the next bracket is either  $w$  or  $1-w+z$ . Minimizing the worst case probability yields  $z = 2w-1$ .  $z$  is positive only if  $x$  is in the larger segment ( $w > 1/2$ ). If  $z$  is “optimal” then so was  $w$ , which suggests that  $z / w = 1-w$ , which in turn gives  $w^2+w-1=0$ , i.e.  $w \approx 0.61803$ .
- Optimal choice of  $x$  is  $(1-w)$  fraction into the larger interval from the middle.

## A Note on Accuracy

- Convergence is linear: count of successive significant figures grows linearly with iterations.
- Do not set tolerance in  $x$  to lower than the square root of your numerical precision.
  - Close to a minimum at  $b$ ,  $f(x) \approx f(b) + \frac{1}{2}f''(b)(x - b)^2$ , because the first derivative vanishes.
  - The second term above is  $\epsilon$  times smaller than the first, i.e.  $\epsilon|f(b)| > \frac{1}{2}f''(b)(x - b)^2$  if  
 $|x - b| < \sqrt{\epsilon |b| \sqrt{2 |f(b)| / \{ b^2 f''(b) \}}}$ , where the final square root is a number of order one for most functions.
  - Unless you know a better estimate for the final square root, apply the limit suggested above.

# One-dimensional Optimization: Fancier Methods

- Parabolic interpolation: try to fit a parabola through  $[a,b,c]$  and jump to its minimum as a guess for  $x$ .
  - Caveats: parabolic interpolation may converge to a maximum, or the fit might not be feasible because of collinearity.
- An ideal scheme can avoid unnecessary evaluations, switch between a robust and slow (e.g. Golden Section) technique and parabolic interpolations as  $f(x)$  permits, and carefully define a stopping criterion.
- Brent's method is good at all the above at the expense of maintaining six function points instead of three and defining robust rules for acceptance of guesses.

## One-dimensional Optimization: Fancier Methods (Cont'd)

- What if we can also use derivatives?
- Keeping the bracketing idea and updating the bracket based on function as opposed to derivative information is more robust.
- You can try to fit a higher order polynomial to function and derivative information.
- You can also try to select the interval to look at based on the derivative at the middle point.
- The latter idea combined with extrapolation to zero of the derivative and robust rules for acceptance of the results as in Brent's method appears to work well.

# Multidimensional Optimization: An Overview

- Methods without using derivatives
  - Downhill simplex
  - Line methods in general
  - Direction set methods (line methods)
- Methods with using derivatives
  - Conjugate gradient methods (line methods)
  - Quasi-Newton methods
  - Levenberg-Marquardt method
- Methods using random numbers
  - Simulated annealing
  - Genetic algorithm
  - Ant colony optimization

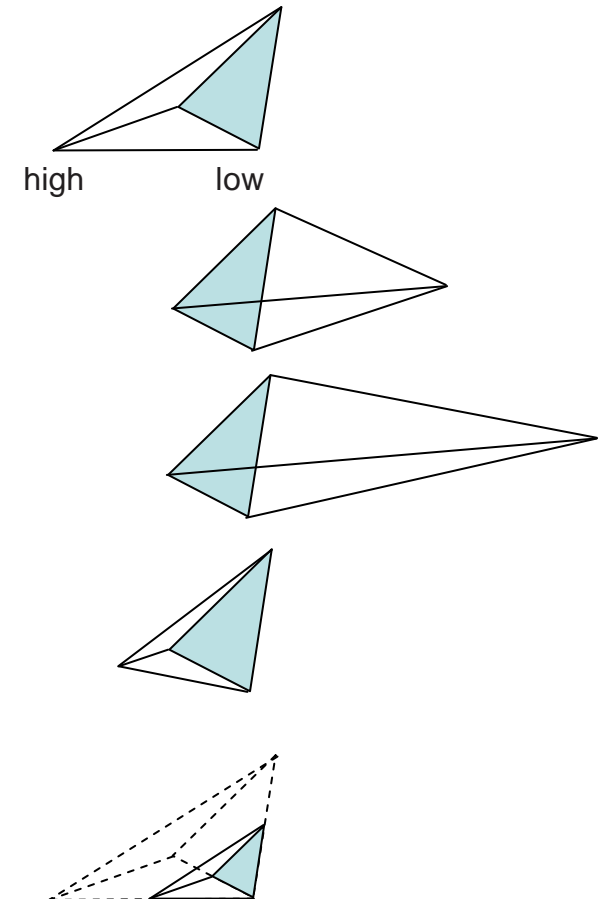


# Downhill Simplex

- Definitely not the best in the number of required function evaluations, but can be used for initial trials. Also, sometimes more robust than other methods.
- A simplex is a geometrical figure consisting of  $N+1$  vertices and all their interconnecting line segments or polygonal faces, etc. in  $N$  dimensions.
- We are now interested in simplexes with a finite inner  $N$ -dimensional volume ( $>0$ ).
- We need an initial simplex when starting the algorithm.
  - e.g.  $\mathbf{P}_0$  is an initial starting point, and  $\mathbf{P}_i = \mathbf{P}_0 + \Delta_i \mathbf{e}_i$ , where  $\mathbf{e}_i$  are the unit vectors and  $\Delta_i$  are some characteristic length scales.
- The one-dimensional bracketing does not work, so we take a slightly different approach to modify the simplex in the iterations of the algorithm.

## Downhill Simplex (Cont'd)

- Given an initial simplex, the following list of moves is considered at each iteration:
  - **reflection:** most of the time just move the point with the largest function value through the opposite face of the simplex
  - **reflection and expansion:** the above combined with increasing the search step size, if reflection yields a significantly better point
  - **contraction:** when in a “valley”, the simplex is contracted in the transverse direction, i.e., when the reflected point is worse than the one with the second largest function value
  - **multiple contraction:** when passing through the eye of a needle it is best to contract around the point with the lowest function value, i.e., when none of the above works



## Downhill Simplex (Cont'd)

- Solution at each iteration is the best point found so far.
- Stopping criteria can be tricky:
  - terminate when vector distance of last move is less than some tolerance limit (not smaller than square root of machine precision)
  - terminate when decrease of function value in the last iteration is less than some tolerance limit (can be approx. machine precision)
  - it is often a good idea to restart the algorithm where it claims to have found a minimum, because an anomalous step might cause the stopping criteria to be triggered

## Line Methods in General

---

- Let us now build on the available one-dimensional routines.
- If we have a starting point  $\mathbf{P}$  and a vector  $\mathbf{n}$  in  $N$  dimensions, then (1) we can use our one-dimensional minimization routine to minimize  $f(\lambda) = f(\mathbf{P} + \lambda\mathbf{n})$ . Thus  $\mathbf{n}$  is the search direction.
- We can then (2) reset our starting point to the minimum found along the search direction and (3) determine a new search direction and continue with step (1) above.
- The various line methods differ by how they choose the search direction.

## Direction Set Methods (Line Methods)

- First, not too bad idea:
  - take a set of directions, e.g. the unit vectors  $\mathbf{e}_i$ ,
  - iterate over the set of directions and minimize along the current direction, then from there along the next direction, and so on until the function stops decreasing.
- Problem: what if the function has a narrow valley at an angle to the unit vectors?
- Let us use conjugate directions,
  - which are directions that do not impact the efficacy of minimization along themselves.
  - If we minimize along some direction  $\mathbf{u}$ , then the gradient of the function must be perpendicular to  $\mathbf{u}$  at the minimum.
  - Using Taylor series:  $f(\mathbf{x}) = f(\mathbf{p}) + \mathbf{b} \mathbf{x} + \frac{1}{2} \mathbf{x} \mathbf{A} \mathbf{x} + \dots$ , where  $\mathbf{b}$  is the gradient at  $\mathbf{p}$  and  $\mathbf{A}$  is the Hessian at  $\mathbf{p}$ .
  - The approximation of the gradient at  $\mathbf{x}$  is  $\mathbf{A} \mathbf{x} + \mathbf{b}$  (derivative with respect to  $\mathbf{x}$ ).
  - The gradient vanishes at  $\mathbf{x}$  that solves  $\mathbf{A} \mathbf{x} = -\mathbf{b}$ , and change in the gradient when moving along  $\mathbf{x}$  is  $\mathbf{A} (\delta \mathbf{x})$ .
  - If we move along direction  $\mathbf{u}$  to a minimum, then a new direction  $\mathbf{v}$  does not spoil our minimization as long as  $0 = \mathbf{u} \mathbf{A} \mathbf{v}$ , i.e., the change in the gradient along  $\mathbf{v}$  is perpendicular to  $\mathbf{u}$  (note that  $\mathbf{u}$  and  $\mathbf{v}$  are conjugate vectors).

## Direction Set Methods (Cont'd)

- If we can find a set of  $N$  linearly independent, mutually conjugate directions, then  $N$  line minimizations will reach exactly the minimum of a quadratic form.
- If our  $f(\mathbf{x})$  is not exactly quadratic, we will be close, and repeated cycles will yield quadratic convergence.
- Powell's method: (1) Initialize the set of directions  $\mathbf{u}_i$  to the basis vectors  $\mathbf{e}_i$ . (2) Save your starting position as  $\mathbf{P}_0$ . (3) For  $i = 0, \dots, N-1$  move  $\mathbf{P}_i$  to the minimum along direction  $\mathbf{u}_i$  and call this point  $\mathbf{P}_{i+1}$ . (5) For  $i = 0, \dots, N-2$  set  $\mathbf{u}_i := \mathbf{u}_{i+1}$ . (6) Set  $\mathbf{u}_{N-1} := \mathbf{P}_N - \mathbf{P}_0$ . (7) Move  $\mathbf{P}_N$  to the minimum along direction  $\mathbf{u}_{N-1}$ , and call this point  $\mathbf{P}_0$ . (8) Repeat from step (3).
- $N$  iterations, that is,  $N(N + 1)$  line minimizations will exactly minimize a quadratic form.
- Problem: throwing away  $\mathbf{u}_0$  in favor of  $\mathbf{u}_{N-1}$  tends to produce linearly dependent directions.
- One (and not the best) solution: rerun step (1) after every  $N$  or  $N+1$  iterations.

## Conjugate Gradient Methods (Line Methods with Derivatives)

- Let us now return to  $f(\mathbf{x}) \approx c + \mathbf{b} \mathbf{x} + \frac{1}{2} \mathbf{x} \mathbf{A} \mathbf{x}$ . Each element of  $\mathbf{b}$  and  $\mathbf{A}$  can affect the location of the minimum, so the information content of this form is  $\sim N^2$ .
- When minimizing via direction sets, we collect this amount of information via  $O(N^2)$  separate line minimizations. If we can easily evaluate the gradient, we get  $N$  pieces of new information and hence  $O(N)$  carefully chosen line minimizations should suffice.
- Note that the gradient may need  $O(N)$  function evaluation time to compute, but there might be repeating computations to take advantage of. Also, each gradient evaluation spares a line minimization, which itself requires possibly many function calls.

## Conjugate Gradient Methods (Cont'd)

---

- Steepest descent performs poorly: takes many steps in a long valley.
- Ideally, we want to proceed along a direction that is chosen to be conjugate to the previous gradient, as well as all previous directions to the extent possible.
- Conjugate gradient methods make use of the technique available for solving sparse linear systems of equations in the context of function minimization.



## Conjugate Gradient Methods (Cont'd)

- Let an arbitrary vector  $\mathbf{g}_0$  be our first auxiliary vector. Let the first direction be  $\mathbf{h}_0 = \mathbf{g}_0$ .
- Let  $\mathbf{g}_{i+1} = \mathbf{g}_i - \lambda_i \mathbf{A} \mathbf{h}_i$ , and let  $\mathbf{h}_{i+1} = \mathbf{g}_{i+1} + \gamma_i \mathbf{h}_i$ , where
$$\lambda_i = \mathbf{g}_i \mathbf{g}_i / (\mathbf{h}_i \mathbf{A} \mathbf{h}_i), \text{ and } \gamma_i = \mathbf{g}_{i+1} \mathbf{g}_{i+1} / (\mathbf{g}_i \mathbf{g}_i) \quad [\text{improvement: } \gamma_i = (\mathbf{g}_{i+1} - \mathbf{g}_i) \mathbf{g}_{i+1} / (\mathbf{g}_i \mathbf{g}_i)].$$
- The generated sequence will satisfy  $\mathbf{g}_i \mathbf{g}_j = 0$ ,  $\mathbf{h}_i \mathbf{A} \mathbf{h}_j = 0$  and  $\mathbf{g}_i \mathbf{h}_j = 0$  for  $j < i$ .
- Problem: as opposed to the case of solving the linear system of equations, here we don't know  $\mathbf{A}$ .
- Solution: let  $\mathbf{g}_i = -\text{grad} f(\mathbf{p}_i)$  and we go along the direction  $\mathbf{h}_i$  to the local minimum of  $f(\mathbf{x})$  at  $\mathbf{p}_{i+1}$ . Let then  $\mathbf{g}_{i+1} = -\text{grad} f(\mathbf{p}_{i+1})$ . This is equivalent to our choice of  $\mathbf{g}_{i+1}$  above, and does not require the knowledge of  $\mathbf{A}$ .

## Conjugate Gradient Methods (Cont'd)

- Described method is called Fletcher-Reeves. Improvement is called Polak-Ribiere: provides smoother transition between iterations and falls back to using local gradient when additional gain is lost.

## Quasi-Newton Methods (Methods with Derivatives)

- Let us now again return to  $f(\mathbf{x}) \approx f(\mathbf{x}_i) + (\mathbf{x} - \mathbf{x}_i) \text{grad } f(\mathbf{x}_i) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_i) \mathbf{A} (\mathbf{x} - \mathbf{x}_i)$ . Recall that we do not know  $\mathbf{A}$ , and so our aim is to collect its information content somehow.
- Basic idea: try to construct an iterative approximation  $\mathbf{H}_i$  to the matrix  $\mathbf{A}^{-1}$  to mimic Newton's minimum search method. That is, making the gradient at  $\mathbf{x}$ ,  $\text{grad } f(\mathbf{x}_i) + \mathbf{A} (\mathbf{x} - \mathbf{x}_i)$ , equal to zero yields the equation  $\mathbf{x} - \mathbf{x}_i = -\mathbf{A}^{-1} \text{grad } f(\mathbf{x}_i)$ , where we use  $\mathbf{H}_i$  to replace  $\mathbf{A}^{-1}$ .
- The name "Quasi"-Newton comes from not using the actual Hessian but the current approximation of its inverse only. It is actually better, because we need to descend, hence we want  $\text{grad } f(\mathbf{x}_i) (\mathbf{x} - \mathbf{x}_i) = -(\mathbf{x} - \mathbf{x}_i) \mathbf{A} (\mathbf{x} - \mathbf{x}_i) < 0$ , which means  $\mathbf{A}$  must be positive definite. There is no guarantee that the Hessian will always be positive definite, but the series  $\mathbf{H}_i$  can be defined so.

## Quasi-Newton Methods (Cont'd)

- Derivation of the iterative approximation  $\mathbf{H}_i$  is somewhat involved, but the two main updating formulas are referred to as Davidon-Fletcher-Powell (DFB) and Broyden-Fletcher-Goldfarb-Shanno (BFGS), the latter being recognized as empirically superior.
- Note that it might be worth using Quasi-Newton Methods with finite difference approximations of the gradient to decrease total computation effort compared to using methods without derivatives.

## Levenberg-Marquardt Method (Methods with Derivatives)

- Let's say now that our  $f(\mathbf{a}) = \chi^2(\mathbf{a}) = \sum [\{y_i - y(x_i|\mathbf{a})\}/\sigma_i]^2$ , i.e., we are looking for a least-square fit. We again use the quadratic approximation  $\chi^2(\mathbf{a}) \approx c + \mathbf{b} \mathbf{a} + \frac{1}{2} \mathbf{a} \mathbf{A} \mathbf{a}$ , which suggests using either  $\mathbf{a}_{\min} = \mathbf{a}_{\text{cur}} + \mathbf{A}^{-1}[-\text{grad } \chi^2(\mathbf{a}_{\text{cur}})]$ , if the quadratic approximation is good enough, or a steepest descent  $\mathbf{a}_{\text{next}} = \mathbf{a}_{\text{cur}} - \text{constant} * \text{grad } \chi^2(\mathbf{a}_{\text{cur}})$ , if it isn't.

- Here we do know the form of the Hessian:

$$\alpha_{kl} = \partial^2 \chi^2 / (\partial a_k \partial a_l) = 2 \sum 1/\sigma_i^2 [ \partial y(x_i|\mathbf{a}) / \partial a_k * \partial y(x_i|\mathbf{a}) / \partial a_l - \{y_i - y(x_i|\mathbf{a})\} * \partial^2 y(x_i|\mathbf{a}) / (\partial a_l \partial a_k) ]$$

- Now, we neglect the second partial derivatives in the above expression, because
  - they are often small enough to be negligible in practice;
  - the term multiplying them should be a random, uncorrelated measurement error in a successful model, which can have either sign, so they will probably cancel;
  - they might lead to destabilizing in the presence of outliers.

## Levenberg-Marquardt Method (Cont'd)

- Levenberg-Marquardt solves the equation  $\sum \alpha_{kl} \Delta a_l = \beta_k$  (r.h.s. is the gradient) by replacing  $\alpha_{kl}$  with  $z_{kl}$ , where  $z_{jj} = \alpha_{kl} (1 + \lambda)$ , and  $z_{jk} = \alpha_{jk}$ . This will ensure a smooth scaling between the two extremes of using the inverse Hessian and using the steepest descent.
- Outline of algorithm is then: (1) given an initial guess  $\mathbf{a}$ , and an initial setting of  $\lambda$  ( $=0.001$ ), (2) solve the above system of equations for  $\Delta \mathbf{a}$ . (3) If  $\chi^2(\mathbf{a} + \Delta \mathbf{a}) \geq \chi^2(\mathbf{a})$ , then increase  $\lambda$  by a significant factor, otherwise decrease  $\lambda$  by a significant factor. (4) Continue with step (2) after setting  $\mathbf{a} := \mathbf{a} + \Delta \mathbf{a}$ .
- The stopping condition can be tricky. Decrease in  $\chi^2$  less than 1 is often not significant statistically, but 0.001 will put you on the safe side. Method tends to wander around near the minimum in a flat valley.

# Simulated Annealing

- Now think about  $f(\mathbf{x})$  as a function with multiple local minima.
- Apply the analogy of liquids freezing and crystallizing from thermodynamics.
  - At high temperatures, the molecules move freely.
  - At lower temperatures this mobility is gradually lost.
  - If cooling is slow, the molecules can arrange in an ordered fashion leading to (or near to) the state of the lowest energy.
  - If cooling is fast, the result is a polycrystalline or amorphous state with higher energy.
  - The key is to keep cooling slow to allow for atoms to redistribute.
  - A state of energy  $E$  is occupied with probability  $\sim \exp(-E/kT)$  at temperature  $T$ .  $k$  is Boltzmann's constant.

# Simulated Annealing (Cont'd)

- Outline of the Metropolis algorithm
  - Define an initial temperature. (How high?)
  - Generate a random move from state  $\mathbf{x}$  to  $\mathbf{x}+\Delta\mathbf{x}$ . (What moves to pick from? How to choose?)
  - Accept new state with probability  $\min(\exp\{-[f(\mathbf{x}+\Delta\mathbf{x})-f(\mathbf{x})]/kT\}, 1)$ .
  - Decrease T. (How fast? Repeats at the same temperature?)



# Genetic Algorithm

- Think about  $-f(\mathbf{x})$  as a measure of success of an individual in a population (fitness function).

Think about  $\mathbf{x}$  as the genome of the individual.

- Define the following rules of evolution
  - Set up an initial population. (What size? Which individuals?)
  - Select the part of the current population for reproduction. (Randomized based on fitness.)
  - Apply crossover and mutation operators to selected subpopulation. (How to define operators? More than two parents?)
  - Evaluate termination criteria (Number of generations? Average fitness?) and continue with selection step, if needed.
- The details are highly problem-specific.

# Ant Colony Optimization

- How do ants search for food? How do they form trails connecting the colony with the food source?
- To illustrate the concept, think about  $x$  as a Hamiltonian cycle in a given graph with given non-negative edge weights. Let  $f(x)$  be the sum of the weights of edges in the cycle. Minimizing  $f(x)$  now amounts to solving the traveling salesman problem.

## Ant Colony Optimization (Cont'd)

- ACO solution to traveling salesman problem:
  - Let's imagine a number of ants that walk on the graph.
  - We define the following rules for one iteration:
    - Each ant must visit each node exactly once.
    - The smaller the weight and the stronger the pheromone trail the more likely that an ant chooses a particular edge as a next step on its cycle.
  - We iterate over the following steps:
    - Each ant visits a Hamiltonian cycle according to the rules above.
    - Each ant lays pheromone trail on the edges it visited in the current iteration. The shorter the Hamiltonian cycle the stronger the trail.
    - All pheromone trails evaporate (weaken).

## Tips&Tricks, Suggestions

---

- Get to know your  $f(\mathbf{x})$  as much as possible to identify qualitative behavior, parameter ranges of interest and signs of numerical errors.
- Try to use as much information about  $f(\mathbf{x})$  as possible, e.g. go for LM if  $f(\mathbf{x})$  is  $\chi^2$ .
- Try to match the scales of function parameters (coordinates in  $\mathbf{x}$ ).
- Try changing your initial guess.
- When working on a constrained problem, hide the constraints from the optimizer by applying variable transforms.
  - e.g.  $f(\exp(y))$  lets the optimizer work in the range  $(-\infty, \infty)$  even if the domain of  $f(x)$  is  $\mathbb{R}^+$ .
- Recall the shape of objective function value plotted against iterations taken in a random optimization method.
- There is not any single best recipe, so experiment with different approaches.

# References

---

Most of this presentation is based on

Press-Tseukolsky-Vetterling-Flannery: Numerical Recipes, 3<sup>rd</sup> ed,

a highly recommended reference.

# Q & A



# References

This material was prepared by sales, trading, banking or other non-research personnel of one of the following: Morgan Stanley & Co. Incorporated, Morgan Stanley & Co. International Limited, Morgan Stanley Japan Limited, Morgan Stanley Capital Group Inc. and/or Morgan Stanley Dean Witter Asia Limited (together with their affiliates, hereinafter "Morgan Stanley"). Unless otherwise indicated, these views (if any) are the author's and may differ from those of the Morgan Stanley fixed income or equity research department or others in the firm.

This material has been prepared for information purposes only and is not a solicitation of any offer to buy or sell any security, commodity or instrument or related derivative (hereinafter "instrument") or to participate in any trading strategy. Any such offer would be made only after a prospective participant had completed its own independent investigation of the instrument or trading strategy and received all information it required to make its own investment decision, including, where applicable, a review of any prospectus, prospectus supplement, offering circular or memorandum describing such instrument or trading strategy. That information would supersede this material and contain information not contained herein and to which prospective participants are referred. If this material is being distributed in connection with or in advance of the issuance of asset backed securities, information herein regarding any assets backing any such securities supersedes all prior information regarding such assets. We have no obligation to tell you when information herein is stale or may change. We make no representation or warranty with respect to the accuracy or completeness of this material, nor are we obligated to provide updated information on the instruments mentioned herein.

This material may have been prepared by or in conjunction with Morgan Stanley trading desks that may deal as principal in or own or act as market maker or liquidity provider for the instruments mentioned herein. Where you provide us with information relating to your order or proposed transaction ("Information"), we may use that Information to facilitate the execution of your orders or transactions, in managing our market making, other counterparty facilitation activities or otherwise in carrying out our legitimate business (which may include, but is not limited to, hedging a risk or otherwise limiting the risks to which we are exposed). Counterparty facilitation activities may include, without limitation, us taking a principal position in relation to providing counterparties with quotes or as part of the ongoing management of inventories used to facilitate counterparties. Trading desk materials are not independent of the proprietary interests of Morgan Stanley, which may conflict with your interests. Morgan Stanley may also perform or seek to perform investment banking services for the issuers of instruments mentioned herein.

Any securities referred to in this material may not have been registered under the U.S. Securities Act of 1933, as amended, and, if not, may not be offered or sold absent an exemption therefrom. In relation to any member state of the European Economic Area, a prospectus may not have been published pursuant to measures implementing the Prospectus Directive (2003/71/EC) and any securities referred to herein may not be offered in circumstances that would require such publication. Recipients are required to comply with any legal or contractual restrictions on their purchase, holding, sale, exercise of rights or performance of obligations under any instrument or otherwise applicable to any transaction.

The securities, commodities or other instruments (or related derivatives) discussed in this material may not be suitable for all investors. This material has been prepared and issued by Morgan Stanley for distribution to market professionals and institutional investor clients only. Other recipients should seek independent investment advice prior to making any investment decision based on this material. This material does not provide individually tailored investment advice or offer tax, regulatory, accounting or legal advice. Prior to entering into any proposed transaction, recipients should determine, in consultation with their own investment, legal, tax, regulatory and accounting advisors, the economic risks and merits, as well as the legal, tax, regulatory and accounting characteristics and consequences, of the transaction. You should consider this material as only a single factor in making an investment decision.

Options are not for everyone. Before purchasing or writing options, investors should understand the nature and extent of their rights and obligations and be aware of the risks involved, including the risks pertaining to the business and financial condition of the issuer and the underlying instrument. A secondary market may not exist for certain of these instruments. For Morgan Stanley customers who are purchasing or writing exchange-traded options, please review the publication 'Characteristics and Risks of Standardized Options,' which is available from your account representative.

The value of and income from investments may vary because of changes in interest rates, foreign exchange rates, default rates, prepayment rates, securities, prices of instruments, market indexes, operational or financial conditions of companies or other factors. There may be time limitations on the exercise of options or other rights in instruments (or related derivatives) transactions. Past performance is not necessarily a guide to future performance. Estimates of future performance are based on assumptions that may not be realized. Actual events may differ from those assumed and changes to any assumptions may have a material impact on any projections or estimates. Other events not taken into account may occur and may significantly affect the projections or estimates. Certain assumptions may have been made for modeling purposes only to simplify the presentation and/or calculation of any projections or estimates, and Morgan Stanley does not represent that any such assumptions will reflect actual future events. Accordingly, there can be no assurance that estimated returns or projections will be realized or that actual returns or performance results will not materially differ from those estimated herein. Some of the information contained in this document may be aggregated data of transactions executed by Morgan Stanley that has been compiled so as not to identify the underlying transactions of any particular customer.

Notwithstanding anything herein to the contrary, Morgan Stanley and each recipient hereof agree that they (and their employees, representatives, and other agents) may disclose to any and all persons, without limitation of any kind from the commencement of discussions, the U.S. federal and state income tax treatment and tax structure of the transaction and all materials of any kind (including opinions or other tax analyses) that are provided to it relating to the tax treatment and tax structure. For this purpose, "tax structure" is limited to facts relevant to the U.S. federal and state income tax treatment of the transaction and does not include information relating to the identity of the parties, their affiliates, agents or advisors. In the U.K., this communication is directed in the U.K. to those persons who are market counterparties or intermediate customers (as defined in the U.K. Financial Services Authority's rules). In Japan, this communication is directed to the Qualified Institutional Investors as defined under the Securities Exchange Law of Japan and its ordinances thereunder. For additional information, research reports and important disclosures see <https://secure.ms.com/servlet/clis>. The trademarks and service marks contained herein are the property of their respective owners. Third-party data providers make no warranties or representations of any kind relating to the accuracy, completeness, or timeliness of the data they provide and shall not have liability for any damages of any kind relating to such data.

This material may not be sold or redistributed without the prior written consent of Morgan Stanley.

Last Page footer:  
© 2006 Morgan Stanley